

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

65/162

158761	GCACTTTGGG	AGGCAGAGAC	AGGCAGATCA	CCTGAGGTCA	GGGGTTTGAG	ACCAGCCTGG
158821	CCAACATTGT	GAAACCCCAT	CTCTACTAAA	AATACAAAAA	TCAGCCAGGC	ATAGTGGTGC
158881	GTACCTGTAA	TCCCACGCTA	CCCGGGAGGC	TGAGGCGCTG	GAACCCAGGA	GGCAGAGGCT
158941	GCACTGAGCT	GAGATTGCGG	CACTGCAAGC	CAGCCTGGGT	AACAGCGAGA	CTCCATCTCA
159001	AAAAAAATT	TGAAAAAGA	AAAATTTTAA	TAAACAGTGT	TTAAGAGGGG	AGAAATATTT
159061	AGTTAAAAGA	TAAGCCCAT	TAAGAAATAG	TTTCACTTGA	CCCGGAAGGC	GGAGCTTGCA
159121	GTGAGCCGAG	ATCGCACCAC	TGCACTCCAG	CCTGGGCGAC	AGAGCGAGAC	TCTGTCTCAA
159181	AAAAAAAAAA	AAAGAAAGAA	AGAAAGAAAG	AAATAGTTTC	ACTTGAACCA	TATTATGATT
159241	CCTTCTGTAA	AAGATGAGAG	TAGGCAAAAT	GACTCAGTGA	AATCCCAGCA	AACTTACAC
159301	AAAGTCTTGT	TCTTCCTTCC	TGTCATCTGT	ATAGGATGAA	ATACAGATG	CTTTTGGGT
159361	TTGTTGTTGT	TTGTTGTTGT	GTATTTGAGG	GGAACACAGG	TCTATAATTC	CTTTTCTGAA
159421	ATCCCTGGAA	CAAAATGGGC	TTTGCCATTC	AAATTAGTTT	AGAAGTTATA	AAGGCAAAAA
159481	AATGCATATA	CTCTAAAGTT	CAACCCCATC	ATGGCCTAAG	GCAGAGCCCT	GTAATCAAAT
159541	TCATCAATAT	ATCTGCAGCA	AAACATTTAT	TCAAATTAAG	TGGGATAAAT	AAAGACTTTT
159601	AAATAGTCTC	ATCTCAGTGC	CGTTCAGGGT	TGGCCACTGT	GGAAGACAGA	CTCAAGGGTG
159661	GCCTTCTATG	ATTCTGCCT	CTTGGTGTTC	ACACCCTCGT	AAAATTCCTT	GTCTTTGAGT
159721	GTGAGCAGGG	CTTATGAATT	GCTTCTGACC	AATAGGATAT	GGCAAAGATG	ATGGGATATA
159781	ATTTCTATGA	TTACGTTTCA	TTATGTAAGA	CTCCATCTTG	CTGGCAGATT	TTCTCTAAAG
159841	AGTCTGTCTC	CTGAGCTCTC	TCTGAAGAAA	TAACTGGCCA	TGTTAGAAGC	CCATGTGCAA
159901	AGAGCTGAGG	GGTGGCCTGT	AGAAGCTGTG	GGCAACCTCC	AGCCAACAGC	CAGAAATAAC
159961	CAGGGCCAAA	GTCCTGCAAC	CATCAGGAAA	GAAATTCTGC	CTGCTACCTC	AGTGAGCTTG
160021	GAAGTGGATT	CTTCCTTAGC	CTAGCCTCCA	GATAAGAACA	CAGCCTGACC	AACACCTTAA
160081	CTGCAGCCTT	ATCAGACCCT	AAGCAGCAGG	CCCAACTAAG	CTGTGCCCAG	ATTCTGAAC
160141	CACAAAAATT	GAGATAACAT	ATCAGTGTTG	TATTAAGGTT	CTAAATTATG	GTAATTTGTT
160201	TGTACTAATA	GATAACTAAT	ATAACCACCA	AATCATTTCA	GGTTAGGCCA	GATTTTTGTA
160261	GCCAAATGAA	TCATGATAAA	ACTTTCATT	TTCAGGGGTT	TTTTTGATTT	TGTACTTACG
160321	GATACAAATT	TGTGAAAGTA	TAGTCAGCAC	TGATTTAAAA	AATCAAGGGA	GCAGGAAACT
160381	CAGTAAATGG	TTCTAACATT	TTGGAACTCG	TAAATTGGTT	GTAACATTTG	GTACTGTGT
160441	TATCTAAGTC	AAGTTCCTAA	AATATGTGAA	TGATAGGTTA	TCATACTCAC	CTACTTTTCT
160501	TGCATTGCTC	TAAGAGTTGG	CTGAGCTATT	GATAATAAAC	ACTATGATCA	GATCTAATAC
160561	CATGATGTGC	TATTATGATC	ATGTGTCAGT	CACAGGGCTA	AGCACTTTGT	ACATGTTGAT
160621	GCATTTAATT	TTGATGATAA	CTCAATGAAG	TAGGAGCTGT	TAATATTTTC	ATTTTTCAGA
160681	GGGGGAAACC	AAGTCACTTG	GAGTAACATG	GCTAATAAGT	GAAAGAATAA	GAATTTGAAA
160741	GGTTTGACAC	GATAACCAGA	ATGCAATGCT	CATCACATTC	ACTGAGCAGT	GAATCATACT
160801	AACTAGAGAA	AGTATGAAAG	CTCTACTGAA	ATTAACTAAA	CAACCTCTCT	GGCTGTGAGC
160861	CTGCCAAGGG	ACAGGTGGTA	AACCTGGTTA	CTGCATAAGG	CCCCTTCTAT	CCACAGTATT
160921	CAGGAATTCT	TTAGTGAACA	TACCTTGATG	ACTCCTTAAC	ATTTTCTTCA	CATCGAAGTA
160981	AAGCTTGGAA	ACATTGCACA	TAGTATGAAG	TTCCAAGGAG	ACAGCCTCTG	ATGTTTCCAG
161041	CTTCACAGCC	CAACTCCTAG	AATAAGCAGA	GGCGAGAGAT	TTCTTCAGAG	GTGCATTCCA
161101	TTCATTTCTA	TATACGCACA	CCCCTCCCCT	CCTGCATTCA	AACAGGACTT	ACCTGCTCAA
161161	AGTGTCAATC	ACATTCTATA	AAGAAACAAA	AAGAAAAGGT	GAGCATGGGA	ACATCGGTAT
161221	TTCATGGGGC	TTGTCATGCA	GGGCTATTCT	TCTTTGCTTT	ACCCGAAGAA	GTAAAGAGAG
161281	TTACCCTAGT	CTTAGTCTTA	GATATTGATG	GATACTCAA	CAAAGTAATT	CCCACCAGTC
161341	TTAGGTATTG	ATGGATACCC	AGATGGAATA	ATTCCTACCA	GCTTCTGGGA	GATTCAGCAT
161401	GGCAGGATGT	TTATCAACAT	TTGCATCTAT	TCTCATCCTT	GCTGAAGTCT	GAGGGCCAGG
161461	AGCTTTGTCC	ATGCTCCCTC	TGTAAGGACT	AGCTTTTGGT	GATCGGATTT	CCTTCACAGT
161521	GAGCCCAGAT	TAGAGAACAC	TTATCATAAA	GGTCCTTAGT	GGTGAATCTG	TGCACAGCCC
161581	TGAGACTGGG	CCACTGCCAC	TAAGATGGTG	GTAGCAGGTA	TCACACAGTG	GTAAAGCAAT
161641	CATGCTATAC	ACTCAGCCTT	ACAGTATAGT	CACCAATCCT	GTTAGTTAGA	ACCAGAATTA
161701	ATGGCTCCAG	ATGTTTATCT	TCCTACAGAT	AAAGCTGTAG	ATTGTACCAT	AACAGCTCTG
161761	GAGCAAGGGT	TCTACAAGCA	AATCAGGGAA	AAGGTTATCA	CTCATTTTGG	CTGCCCCACT
161821	TCATCACCCA	TCAGTCACCT	AGTGAGTAGT	TTCAGGAGAG	AGTCAACAAC	CAGGGTTCTC
161881	TGCACATGGG	CCAAGGAGGC	AAACAGTGGT	AAATGTTATC	CCGTGGTTTC	ATTTGGCCAA
161941	GCTGTGTTCC	CTCAGAAGTT	TATTTTCTA	ATTGACATAA	AGGTACCCTA	TAAATTAGTG

Figure 8 (Page 50 of 73)

66/162

162001	AAGGCCAGCC	TGATGGCACT	GATGTACATC	TAAAAGAAAC	ATTACTTTAT	CTTCCCATGC
162061	TTCCTTACCA	TTCTCCTTTA	ATAGCACTAT	AACATACCTT	TTTTCCCTAC	TCCAAGTACA
162121	CAGCCTCACC	TGCAGCAATT	TCTGGGCTGA	GCCCTGACAT	TTTTCTCTCA	GTTCCAGGAT
162181	GTGGCTCTTG	AGTTCATTGC	TCTTCAGCCC	CAGACCAGCC	TCATAGTCCC	TCAGTCTACT
162241	CAGAGTCTGT	TGTTCTTCTT	TCTCCAGCCT	CCAGAGATAA	GACTTCTCTT	CCTCATGTAG
162301	GAAACACTGG	AGATTCTTAA	AGTCAGACCG	GATTTTTTGT	CTCTGAATCT	GTACCTTCTC
162361	CTGGAGTCAA	GAAAGTATGG	TCAAAAGGTG	GAAGTAAACC	AAATGTCCAT	CTATGGATGA
162421	ATGGATAAAC	AAGAATGAAA	GTCTGACACA	CGCTACTACA	TGACAAGCCT	TGAAGACATT
162481	CAAGCAAAAT	AAGCCAGAAA	CAAAAGGGCA	AATATTGTAA	GACTTTGCTT	ATACAAGGCA
162541	TCTGGAGTAG	TTAAGTTCAT	AGAGACAGAA	AGTAAATAG	TGGTTACAAG	GTGTTGGCAA
162601	GACCAGAAAA	TGGACAGTTA	TTGTTTAAATG	GGTAGTGAGT	TTCAGTTTAT	AAGATGAAAG
162661	ATGAAACTGA	GTTGCAGTTT	GGAGATGGGA	ATGGTGATGG	TTGCACAACA	ATGTAACAAT
162721	GTAAAAGCAC	TTAATTCTAC	TGAACTATAT	ACTTAAAGT	GGTTAAATGC	TTAAGTGTTA
162781	TATATATTTT	CACACAAACA	CACACACACA	CACAATCAGC	CACTGGGACA	TTATTTTCTC
162841	ATGAGTCACT	GAAGCTGGAA	GAATGTCCCC	AGTTTCCTGC	TGCAGAGTCA	TGTGTGGGAG
162901	GCAGGCACTC	AGATGTGGAA	GAGGTTGCCT	CAGATTCCTT	ATAGTCACCC	AATTAATTTT
162961	CTTGTTCTTC	AGCCAAGACA	CAGGAGAAAG	CTGGGTTAGG	AGTGCTAGAT	AATTTAATTG
163021	TGAAACTAGG	GCCAAGTTCA	AACACTTTAT	CAGTTACAAG	GATAAAAAGA	GGTTTTTACT
163081	TATGATTTAA	GAAGTTAGAT	TTCTGAGTTG	GAGCGATTTT	CTTGAAGTAA	AAGCTTATAA
163141	TGAACATCAC	CCAGACTGGA	TTTTAAGACA	ACCAGGCTGG	TAAGAGGGTC	CATAATTCTT
163201	GGCAGGGGGA	GCTTTGAGTG	TGACAGGCAT	TTATTATGGT	TAAGTGAGAA	ATACTGTTCT
163261	ACTACCCTAG	GGTCATCTTA	AGCATTCCTA	TGTGTAAGAC	TGACAGAAAT	CAAGTGAAC
163321	TCTCATCTGA	GGAGATGTAA	AGTTGCAATT	TCCATTAGTG	CTGTCTAAAT	TAATGCAGTG
163381	GGAGTGTGTA	TTCAGGGCAA	TTTGAATCTA	TGTTCTTGGA	TTGCAGTCTT	CAAACCTGGC
163441	CCAAATAAAC	TCTCTACTTA	TCTTAAAAAA	ATAAAAATTA	AAAAATAAAA	ATAAATTCAT
163501	ACAGTGTTTT	GATGACTATG	ATATAGAAGA	AGGGTCTTTG	ACTTAGGATG	AGGTGGAATT
163561	TTTGTGTAGG	AGACAGGTGC	AGCTTTAACT	CTTGATAGTA	CGGGTTTTCA	TATATGTTAG
163621	TTACAATCAA	GGCTTCCCTC	ATTGCCCAAG	ATCCTAGAAA	TGGGGGAAGT	AAGAGTGTAC
163681	TCAGGAGCTC	AAGAGCAACA	TCCACAAACA	AAGATCAGGG	TAGAGGTTAG	AGAGGACTCC
163741	TGAAAGAGAG	AAAATTGGTA	ATCAGCTTGT	GGGATTTTAC	TGCAAGCTAG	TGAATTATAT
163801	AAATATAAAG	ATTGGTGCAA	AAGTAATTGT	GGTTTTTGCC	TTTACTTTAA	TGGCAAAGAC
163861	CGCAATTACT	TTTGCACAAA	CCTAAATATT	TCCATAAAAG	AATGTGGCTC	TGATAATGTG
163921	GAGGTTAGTC	AGCCACGGAA	ATAATCTGAA	AGTTTGTAAGT	TGCAAGTGTG	TAGGTTGTTG
163981	CATTACTTGT	GATGTACTTA	TAAATCAAGT	ATAGGCCGGG	TGCAGTGGCT	CACGCCTGTA
164041	ATCCCAGCAC	TTTGGGAGGC	TGAGGTGGGT	GAATCACGAG	GTCAGGAGAT	CAAGACCATC
164101	CTGGCCAACA	TGGTGAAACC	CCGCTCTTAC	TAAAATACAA	AAAATTAGCC	AGGCATGGTA
164161	GCACATGCCT	GTAATCCCAG	CTACTCAAGA	GGCTGAGGCA	GGGGAATTGC	TTGAACCCGG
164221	GAGGTGGACA	TTGCAGTGAG	CTGAGATCGC	ACCACTACAC	TCCAGCAAGA	CTCCATCTCA
164281	AAAAATAGTA	ATAATTTAAA	AATAAATAAA	TAAATAAAGT	ATATTTCTTT	CATCAGCTTC
164341	ATGAGCTAGA	GTAGTATGAA	TTTCAATCTG	GAGTGATCCT	GTTTTCTAAG	TGTTCAACAA
164401	GCTTGGTTTC	TGTACCTGTA	AAGTTGAGAG	CCAGATGCTC	CACTGTGGTA	AAAGTGCCAG
164461	GGTAATGAGT	TGAGGCCCTG	AAACCAGGTT	TATTTTGACG	TATTTAAAGT	TTGAGACCCA
164521	CTCGATGCTT	TTTCTAGGTA	AATAGTCATA	CTAATTCCTG	TTCTTCTGAC	TGAAGTATCA
164581	GGAATCCAG	CCAACACAG	TTTAAAGATG	GAAAGATTGG	TGCTAAATAC	TCATGGATGT
164641	AAACCTGGAA	CCAGGGGCAT	AAGTACAAAT	AATGGTTTCT	TCCTTGGGTT	TCATTTTTTC
164701	AATCTGGTTT	AGTGAGAATA	AATCCTCATT	GTGCTTTTCC	TCAATCATCC	CCTATGCCTA
164761	AGCTCTAGAA	TGGAAAATAG	CTTGAGATCA	ATGAAGTCAG	ATTCTTACTT	TCCATTTAGT
164821	TATTCGCATT	GCTGTGGACA	GCTTCTGCTC	CGTACATCTG	TCTTCAAGTT	GCTTCAGTTT
164881	TGTCACAGCT	TTCTGGAGCT	TTTCCTGAAG	GAAAAATTG	ATAAGTGAAG	CCTATTCAAT
164941	TTGACTCTTC	ATTAGGGACC	TAGGGGGGAA	CCCAATCTTC	TAAGATATAT	TTGAATAATA
165001	GTGAATATTT	ATAGAGTCCT	CATTGTTTTT	TGCTAGAGAG	CATGCTAAAG	GCTATATGTG
165061	CAGGAACATA	CTGATCCCTT	TGGCAACCTT	GAATAGTTGG	TAGGATTTTA	AACTTCATTT
165121	CTGTGCTGTA	GAAAATGAGA	CTAAGAAAGG	GGTAAATAAA	CTTGCCCAAA	GGGCTATGAC
165181	TGCCAGGTGG	TGGAGCAACA	ATTGCAATCT	CATCTGCTGA	CCCAGAGCCT	GAGCTATGTC

Figure 8 (Page 51 of 73)

SUBSTITUTE SHEET (RULE 26)

67/162

165241	CACCACTAGA	GTCCTGCCAG	GAAAAAGTTG	GATATAGAAC	AAGGTAATCA	TCATCTAAAA
165301	GATTTTGTA	AACAACATGC	TGAACCAAGC	AAAACCAATA	CCAGTGTTTG	GCACACATGA
165361	AATTTTGTGT	CTTATGAGTC	AGGAAAAATC	AGGATGCCAG	CTGGTTATTA	GAAACAGTTC
165421	ATGGAAGAGG	GGAATCTCTG	TATCTTTTGA	ACAATGGTAT	CATGAATCCA	ATTTAAATG
165481	ATTTAGTATT	CATGTCAAGC	TTTTAGCTTA	TTCTTCAAAA	CAGTTTCTCA	TATTTCTATT
165541	GAAAGTGATT	TGAAGCTGAC	CCAAATTGCT	AATTGTAGTC	AATGCTGAAA	GAATTGTCTC
165601	CTGTCTCTG	TAAACCCAAC	AAGTATACTC	ATTCATTCTC	GAGTGTCTC	AGGAAAAGGT
165661	TCTATGTAAC	TGTTT TAGCA	AAAGATGACA	TTGTCCTTAC	TATATGCCAA	GTGCTATTCT
165721	ATGCATTCTA	TATTTTAATG	TCCTCAAAGC	TTATAACCAC	CTCCTGTGTA	TGTGTTTTAG
165781	GGAGGGAGGA	CACTGCTATT	ATCCCCATTT	ACAGATGGAG	AAACCAAGGT	GTGAAGACAT
165841	TAAGTAACGT	GCCCCAAAATT	GCCCCATCTAG	TAAGTGACAA	AACTCAATTT	CAACATAAGC
165901	TGGTTCCTTT	TCTTACTACT	TGGTGAAAAA	GTAATTCAAA	TGGGAATATG	ATCATCGCAG
165961	TTATTAGCTG	CTCCATGGAG	TTTAAGGAAG	AGCTGCCATG	AGCTGAGTGG	TGGTCATGAT
166021	TGACATGTCC	TTAGAAGGAC	TTAGAGCCTT	CATACAAGAC	CACCTCTGCC	TCATGGAGGA
166081	CAGAATAAGG	AGCCTGACAC	TGGAGACAAC	ATTTCTCTCA	AATTTAGGCA	GGACAGAGAA
166141	GGAAAAAGGA	CATCAGGACT	ATGCCCATTC	CTCCATGCTG	CCAACAGCAA	AGTCCCACCT
166201	TCCTTAATAT	GCTTCTGGC	AAGAAATCTG	GATGGTACAC	AAAACCTCTC	CCTCTGCTTC
166261	ACCTTCCACA	ACCAAGCATT	TCCAAATCTT	TGACTCTTCT	TCCTGAATCG	TGCTTAAAAA
166321	CTGCCCTCTC	CTCCCTTTCT	TATACGGATA	GTTTGAATTT	TACTCCTTGA	TATTCCTTTT
166381	ATCATAGACA	TGCCACAGTA	GCTGGGCACA	GTGGTTCATG	CCTCTAATCC	CAGCATTTTG
166441	GGAGGCTGAG	ATGGGAGGGA	GACCAGGGGT	TTGAGGCCAG	TATAAGCAAG	AAAGGCAGAC
166501	CATGTCTCTA	CAAAAAATAA	AAAAATTATC	CAGGTATGGT	GGGGCATCCC	TGTAGTCCTA
166561	GCTACTTGGG	AGGCTGAGGT	GGGAGGATTG	CTTGAGCCCC	AGAAGGTTGA	GGCTGCAGTG
166621	AGCCGAGATT	GCACCATTGT	ACTCCAACCT	GGGATACAGA	GCAAGACCCT	ACCTCAGGAA
166681	AAAAAAAAAA	AAAAAAAAAA	AAAAGTAGAG	GTACCAGAGT	GATATTTTCA	ATGTCACCTA
166741	CCCTTCATTC	CCCAAATGAA	AATCCCCCAA	TAGGTGTTCA	ATTTTACGT	GTCCTTCAGG
166801	AGTTACTTCT	AAGATGAACC	ACTCTCTACC	CTAAATGTCC	CTCCCCACCA	CCAAAACCAG
166861	GGACCTCCAG	GCAGACATTT	TTGATGGTTT	GTTTCTTTA	CTAGACTGTA	GATACCTAAA
166921	AGGTGATGGG	TCTTCTTCC	CTGTTTTCAG	GCCCTACTGC	ATGGCTTTAC	ATATTGTGGT
166981	TTTTCAAATG	ATATTCATGG	TGTGAAACAA	GAAAAAATGC	GGGTGTTTGG	TTTGAGAACA
167041	ACCTGTTCTA	AAGCAAAAAG	AAATTCATCA	TAACACAAAT	GGATAGAGAT	AAGAGTCCAA
167101	CCATCCCATT	GAAGGTCAGG	ATGGACAGTC	TAGATAATTG	AGCAAGAAAT	CATCATAAAC
167161	TATTTTTTCAG	AAGAATGACA	TGATGAAAGC	TGTATTTCCA	AGTCATAATG	TTAGGTTTCA
167221	AGTTAAATCA	TCTCAGCTCC	TGGGGAGCAG	GATAAGACTT	GGTACTTACC	AAAGCTCCCC
167281	GGCCACACAC	CTCACCTTGT	AGCCCTGGCA	TACGTCTTCA	ACAAGAGCTG	TGGTGTGCCC
167341	TTGTGCTGT	GGTGCCCGCT	CACAGCGCCA	GCAGATGAGC	TGCCCTCGT	CTTCGAGAA
167401	CAGGTGGAAC	TGCTCTCCGT	GTTCTTCACA	TGACATTTCT	TGATCCGTCT	CTTTGAGGGC
167461	TTCAATGAGG	CTTCCAGCT	GCTTGTTGGG	TCGGAGGCTA	TCCATATGAA	ATGGAGCCCG
167521	ACACTGGGGA	CAGCAGAATG	TCTCTGCCT	CAGTTGCTTT	TGGCTTGGGT	TTTTAAAGAA
167581	GTCTGTTATA	CACAAGTGGC	AGTAGCTGTG	TCCACAGTTG	ATGCTTACTG	GGTTCGTCAT
167641	CAGGCTCAGG	CAGATGGAGC	AGGTGGCTTC	CTCCATCATC	TTCTTGGTGC	TGGTGGTTGA
167701	GGCCATAGCT	TTATTGAAA	AGCTCCAATA	TTGGCTCTAG	AGATGGAGAT	GAAGCAGCCA
167761	GAATTTTCCA	CCGTGATGAA	AATACACCTC	ACCTGCACCT	CTATGTGATG	AGCTGGCTGC
167821	AACTGACTTC	CATAGGTCTT	GAAGGTTTTT	CTTCCAACCC	CTATTATCTC	ATTTTGTATT
167881	GAAGAAAAGA	GGACCTAAAA	GGAAGAAGTT	GAGGCTGAGG	TTGTTTGGGC	CACGTTTGAG
167941	AACTGCAACC	CAAGTGCAGA	GTTTCAAGTT	GCCCTCATT	GCAAGCAGTT	ACAAGTGGTT
168001	GTTTAGAGGA	AAAAAAGCAG	TTTTAAAGCA	GTTTTAAAGT	TGTTTGCCAA	GAATTTACAT
168061	TAAATAGCA	TAAGCTTTTG	ACTGGCTATA	CATTGTTCTT	TGTATTACAA	ATCTCGGGAA
168121	TATGTAGGTA	ATAGATGAGG	CAGCCAGTCA	GGAACAAAAT	GCTTTTAAAC	ATGGGGTCTT
168181	AACTGAAGAC	CTATACTCCT	GCCTCACTTG	TCTTGATAAA	TTTGCATAC	CTCACATAGC
168241	TCAGACTGCT	CTAAATTATT	TCATTATTTT	TCTTTTCTCA	GTCTTCTAAC	TTTTTTTTTT
168301	TTTTTTAATG	AGACGGAGTC	TCACTCTGTC	ACCCAGGCTG	GAGTGCAGTG	ACGCTATCTC
168361	GGCTCACTGC	ACCTCCGCCT	CCCGGGTTCA	AGCGATTCTC	CTGCCTCAGC	CTCCCCAGTA
168421	GTAGCTGGGT	CTACAGGTGT	GCACCACTAC	GCCCAGCTAA	TTTTTGATT	TTTAGTAGAG

Figure 8 (Page 52 of 73)

SUBSTITUTE SHEET (RULE 26)

68/162

168481	ATGGGGTTTC	ACCATGTTGG	TGCGCTCGAT	CTCTTGACCT	TGTGATCCAC	CCGCCTCAGC
168541	CTCCCAAAGT	GCCAGGATTA	CAGGCATGAG	CCACCGTGCC	CAGCCTCTTT	TTCTTTTCTT
168601	ATAAGACAAG	TTCTCGCTCT	CTTGCCCAGG	CTGTAGTGGA	GGGCAGTGGC	ATGACCACAG
168661	CTCACTGCAG	CCTCGACCTC	CTGGGTTTAA	GCAATCCTCC	TGCCTCACCC	TGGCAGAGTG
168721	GCTGGGACTA	CAGGTATGTG	CCACCATGTC	CAGCTAAAGT	CTTCTCTCCA	GAAAGAAGAA
168781	ATGCATTGGA	ATTTAGAGGA	TACACAAACA	TCTAGCTGTA	TAGCTAATAC	AGTAGCCACT
168841	ATCATGAGTA	GGAATTTAAA	TTTAACTTAA	TAAAAATTAA	AATGAAAAAA	TTCAGTTTTT
168901	CTGTTCCAGT	TGCCACATTT	TGATTGCTTA	ATAGTTGCAT	GTGACTAGTG	GCTACATAAC
168961	AGCCTCAATA	TACAACATTC	TGTTATCACA	GAAAGTTACC	TTGGACCAAG	TGCTGGGAGA
169021	AGCAATGCAG	GCTTCCTCAC	AAAAGCTGTA	AAAGAGAGAA	CTCAGGGAGT	GTGAAACTCT
169081	TTCTATTCT	AGTTAACTTC	AAGAATAATT	GTTACCAGGC	CAGCACGGTG	GCTCACGCCT
169141	GTAATCCTAG	CACTTTGGGA	AGCCGAGGCG	GGCAGATCAC	CTGAGGTCAG	GAGTTTGAGA
169201	CCAGCCTGAC	CAACATGGCA	AAACCTCATC	TCTACTAAAA	ATACAAAAAG	TTAGCTAGAT
169261	GTGGTGGTGC	ACACCTGTAA	TCCCAGCTGC	TCAGGAGGCT	GAGGAAGGAG	AATGACTTGA
169321	GCTCCGGAGG	GGGAGGTTGC	AGTGAGCCCA	GATTACACCA	CTGCACTCCA	GCCTGGGTGA
169381	AAGAGCGAGA	ATCTGTCTTA	AAAAAAAAAA	AAAGAATAAT	TGGTACCAGA	ATTACTCTTT
169441	GTAATTAGTA	GTAACACTTA	TGCAATTGGG	TGATCTGTGA	CAGATTCCAT	TGAAGGAGTA
169501	TGGGGAGCTT	CACCCCAATA	TATGACTCCC	TGGTATAATG	AGTATTTTGA	ATTAAAGGCC
169561	CTTAGAGATC	AGCAGATGCT	GGAGAGACTC	TTTCCCCTAT	CTACATAAAG	ACCAGTCACA
169621	CTAGACAAGA	AGAACAATTG	TTTTCTCTTC	CAACCCCTAT	TATCTCAATT	TGTAECTGAAG
169681	AAAAGAGGAC	TAAGAATGTA	ACCAGACCTA	ATCAGACACT	TTCACAAAAT	AATGTCTGTC
169741	TCTCAGGCTC	ATTCATTTTC	CAAAGAGAAC	CATTTACAAG	TTAAACTCTG	TTCCCTCCATT
169801	CATTCATCCT	CCCAAATATT	CATTTATTCT	CCCTAGTAAT	CATTTACTGC	CCCTCAAAGA
169861	ATTACCTATA	TTCTCCTGAT	ATCACCTTTC	CCCTCTGAAA	TAAATATGTA	TACATGTATA
169921	AACGTTATAC	ATACATATTT	ATACAGTATA	CATACATATT	TATACATACA	TACATATGCA
169981	TACATATTTA	TATTTATGTA	TTTATACATA	AGTATTTTATA	AATAAGGCTA	TATAAGTATC
170041	TACCCCCATT	GGCAGAGGGG	GTAATCACTC	TGTGATTCTA	GCCCATGTAC	TTGTTAATAA
170101	ATTTGTATGC	CTTTTCTCCA	ATTAGCCTGC	CTTTGTGTAG	TCGATTTTTT	AGTGAACCTC
170161	AGAAGGCAAA	GGGGAAGTGT	TCCCTTGGCT	CCTACACCAT	CATGACAATA	AAATTTGACT
170221	CCACCTCGAC	CCCCCCCATC	CCCCACAAAG	AACAACAACC	AACACTGGTT	AATAAGGTCTG
170281	GTTGTTTTTT	GTTGTGTTTT	TTGTTGTTGT	TGTTTTTGCT	TTCAGGAGCA	GAGGTATAAT
170341	AGGCAAAAGA	AAGAGAAAGG	AGAATAGTGA	ATACCTCTTC	TGCAGAGAGG	GGTGCCTAAG
170401	TGGGACTTCC	CTGGCTAATA	ACGTCTTGCT	AGAGACCCAA	CCAGGAGGAT	AATGGAAGCA
170461	ATCAAGGCAA	CCAGAACAAC	CAGAAGAACC	GGTTTATCCT	TTTGTGCCCC	TCTCCCTAAA
170521	CTGAGGGAAT	AAGAATTGGA	AAGAAGGCTG	CAGAGCAGAG	GGTTTGCTCC	TGAGGAGCAG
170581	TTATTTCTAT	GGGATCAGAG	CTCCTGCAGA	ACTGGGGAGT	TTACTTTTAC	TATCTCTTCT
170641	CCAGGACAGG	ACCTATCTCA	AGAGACATGT	TCAGAGTGAT	TGCAACATAA	AGAGTTTGCA
170701	GACCCAAGGA	GGTAGGGAAG	GCAGAAAGAA	GATGGGGGAG	GCCAGGGATA	GGCAACAGAG
170761	GAGTGACCAG	GAGCGAAAAA	GCCTGCCTCT	TCTGAGAACC	TAGCTGGGCT	CTCCCTGTAC
170821	CCCCGATCCC	TCCCCCCC	CCGCCCCCAC	ACCCCTACTC	CTGGGAGCTC	CTCTAGGACA
170881	GGGGCAGAGT	CAGGAGGAAG	TTTGAAGAGT	GCCTAGAATA	AAAAACAGTA	ATTTAACTAC
170941	AATTACCGGG	TAGGCTGTTT	TCCTCTCACA	ATTTGATCAG	TCTCTTGAAG	CCACACAGAA
171001	TTTCTTCTGA	AGACGTGTAT	TCCTTGCCAG	GCTATTTCTT	CCAGTGATAC	ACCAGGCCCC
171061	TCTCTGCTGG	GGTCACTGCT	CTTCTGGGGA	GATGGGGCTC	CCCTCCTTCC	AAGGCTCCAG
171121	GGTTCCTGTC	CTGGGCCCCA	CTCATCTAAG	TTCTGAATCT	TCTGAGATTT	GGTGTAAGT
171181	CTGGTGAAAG	AAAAGAGCAGG	AAAGAGGTGA	GAGCTGTAAA	ACAAAGAAAG	TCCTGACCAT
171241	TTTCAGAGTT	GGAGGGGCCC	TGCTGTACAG	AAATATATTC	CCCACCCAC	TTGCCATCAG
171301	TACACACTCA	CATATCCACT	GAGAAAACCT	TAGCCTGGAC	CTTTTCCGTA	ACCTTCACTG
171361	CTCAGACACT	TACATATTCT	CTGCTAGTCC	CCTCTGTTGC	TGCCACTTCC	TGGGTCAGGA
171421	AGTTAACTCA	GACCGGATTA	AACTGAGAAG	TGAAACTACT	GTGGGAGGCG	GGGCTCATAA
171481	GATTTAGGAG	AAAAGTAGTG	ACGTTGTTCA	TATCATTTGC	ACTCCGCCTC	TCCGGTAAAG
171541	GAGGGGGAAA	CGTAGGAAGA	AAATATCCTT	CTTTTACAGC	AATAAAAAGA	AGGAACCAAT
171601	TAATAACCCCT	GTAACCTATC	ATGTGACCCC	AACACAGAGT	ATCTAAAAAC	AGGAAGCCTG
171661	CAGAGGTTCA	GTTACACAGAC	TCTGATTTGA	GATCTTTCTA	CTTTTGCCAC	CAACTCCCTT

Figure 8 (Page 53 of 73)

69/162

```

171721 GGGAGTCCTT AAGCCTTCCT AGCTGATGTT ACTTCTTTTG CTATTTATGG GTTGCTTGTG
171781 GTTCTATAAC TGCTCTGAAG GGTGTGGTGG AAAAAGGGGT GGTAACAGCA GTAGGACTCA
171841 TTGGCATCAC AAAATTCATC TGAGTCAGCT TTCTATTCTT CTCTGTCCCG TTCTGTGTCT
171901 TGTTTTTCTC CTTGCTGTCC TTCTGCAGGA CTCAGATCTT CTTCAATAGC GAGGGTCAGC
171961 CAGGATAGAA AATGGGAGTC ACTAGTGGCC CAGCAGTGAG TGCCCCCAGC TTAGAGCTGT
172021 TGGGGATCCC TGGGACCATC ACTCTGCTTT GTGCTTTGTG GAGAAAAGGC TGTGGGGTCC
172081 AGGGTCAAGT CCTTAATGAC TTAGCTCCAG CTTCTCCACT TCAAAATGAA AGGAAAAGTA
172141 CTATCACCAC CCGTTAGAAT TATTATTCCA TGGGGAAAAA AGATGGATTA CTATCTCACA
172201 ATAAGAGCTT GTCACATTTA TAAGTCTCAG GTGTAAGAGG CATTTATGAT AACAACTATA
172261 TAAATGCTGG CTTAAGTAGA TGCAGTGGTC CAAGGGAACC AGTAAGGGGA GCTCAGGACA
172321 CAGGTGGGAG GAGAAATTAA ACTTGAATTC TGGGAGCCAC TGGCCTGTCT GGGCCCCTGG
172381 CCTGCCTGCT GACCCTGATA GCCAATGGAA CATGGAGTTT GGCCCGCTG CAATCCCTCT
172441 GGTCCAACCTA CTCAAAATAA AGGCAAGATT GGGAAACACG TTCCTTTCTT CCTATACCAA
172501 GCAGAAGACT CTTACAGCACT GCACCCTCCT GGGTGCTCAC AGAGCCTTCT GTTGTTTTGC
172561 CACCTACGAT TCATCATGCC CTGGCATGAT GGTTCAGAC CCCATGCATA GCATGGGACA
172621 TTCTACTCCT GAGGCAACCA GCACACAGAG AGAGGAGAAA GAATGAGCCC CTGAATCCTT
172681 GGTCCCACGA TGAGTCCTTG CAGATATCTA CAACTTTCAT TGTTGTGGAT GTGACTCTGT
172741 ACCCAGGCAT GGCTCATTCC AGATCTGTCC TATTGTGAGA GGTGTTCAAA CCAGAATGAC
172801 TCCATTTTGA ATGGGGGCTA GGTAAATAA GGCTGAGACC TACTGGGCTG CATTCCCAGG
172861 AAGTTAGGCA TTGTAAGTCA CAGGATGAAA TAGGCAGTTG GCACAAGACA CAGGTCATAA
172921 AGATCTTGCT GATAAACAG GTTGCAAGTAA AGAAGCTGAC CAAAACCCAC CAAAATCAAG
172981 ATGSCAACAA GAGTGGCCTC TAGTCATTCT CATTGCTCAT TATACAGGAA TTATAATGTG
173041 TTAGCAAGTT AGAAGGCATT CCCACCAGCT CCATAGTGGT TTATAAATAC CATGGCGATG
173101 TCAGGAAGCT ACCCTATATA GTCTAAAAG GGGAGGAACG CTTGGTTCTG GGAATTGCCC
173161 ACATCTTCC CAGAAAACAT ATGAATAATC CACTCCTTGT TTAGTACATA ATCAAGAAAT
173221 AACTGTAAAGT ATCTGTATTA GTCCATTTTC ACCTGCTGA TCCAGACATA CTGAGACTG
173281 AGTAATTTAT ACCAGGAAAA AATGTTTCAT GCTCTTACAG TCCCACGTGT CTGGGGAGAC
173341 CTCACAACCA CAGCAGAAGG CAAGGAGGAG CAAGTCAGGT CTTACATGGA TGGCAGCAGG
173411 CAAAGAGCTT GTGCAGGGAA ATTCCTTTCT ATAAAACCAT CAGGTCTCAT GAAACTTATT
173461 GACTATCATG AGAACAGCAG TATAAATTAC TCAGGGAAAG ACCTGCCCCC ATGATTCAAT
173521 TACCTCCAC CAGGTCCCTC CCACAATATG TGGGAATTTA AGATGAGAGT TAGGTGGGGA
173581 CACAGCCAAA CCATATCAGT ATCCTTAGTC CAGAAGCTGA TGCTCTGCC CTAGAGTAGC
173641 CGTTCTTTTA TTCCTTTACT TTCTTGCTTT CACTTTACTG TGTTAGACTG CCCCAAATTC
173701 TTTCTCACAC GAGATCTAAG AACCTTCTCT TAGGGTCTGG GTTGGGACCC CTTTCTGTT
173761 AACACTATCA AAGGATCAGG AAAAGGAAGC TAGTGAATGC TAAAAGGAA ACAAACTACC
173821 ATTACCAATA ATAACAGCAA GACAAAAGCA AAACGGATTG TGACAGCTGT CCCATCTCAC
173881 ACCTGTTTCC CATTGCAGGA AGGAGGGGCT GGTTCATGCA CAGAGTGGCC AATATTAGAA
173941 GCAGAGATGG GGTGCAGATG AGACTTCAGG AATATGTTGA CAAAGGCAGG CCTAGGGAGA
174001 AATCAACCTG AACTATCCCC AAGGAGGAAT GCATTATCTC TAATATGTAA AGTTAGGCTT
174061 GATCCTGTGA TTATGGGATA TAGGAGTCCA AAGACTCACA ATGGGAAGTA GGTCACTAGA
174121 GTCTCCTTCA GAAGCTCTGT ACTGTGTGTT CCCACTGTGG GCAAGAGTCA GCACCTACGT
174181 ATTCTAGAA TGCCTTTCCT CAACTCCTTC AGATTTTGCC TCTCAACTAA CCCTATCCTG
174241 ACCACTTGTT AGCAAGTGT CCCCTCTCT CCTCCAAAC ATTTTCAAAT CTATTTTGT
174301 CCCATGGCAC TTATCACTGA ATATTTTACT AATTTATTTT GTTTAGTGTT TGCTTCCCTC
174361 ATGAGAATGC AAAGGGATGG ATTTTTTTCA ATATTGTTCA CTGATGAATC CCAGTAACTA
174421 GAATATTTCT AAGCATAGTG ATGTGCATTA AATCAAAGAG TAACTTTCTG AATTGCACTA
174481 AACACACATC ACAAGAGGTG TGTGCACATA TGTGCATGAT GCACGTAGTG TGGTGTGGGT
174541 GTTGTGTGGG GTATGTGGTA CTGTGTGTGC TGTGTGTGGT ATGTGATACA TAGTTTGTGT
174601 TAGTGTGATG CATGTGATGT GGTATGTGTG TGCCTGTCCA TACATATTAG GGGTGGCGGG
174661 GATGTTAATA TGTCAAATGG TACTAGAAAG TATCAGAACT CATGGTGCTT ACTGGTTTCC
174721 CAGAGAGCTG CTTCTCTCCC ACCTGTAGGA TATACTGATG GTTTGGACAG AGAAGAAATA
174781 AAAAGAAGGC TGTGACCTAC TGGGCTGAGG AAATAAAAAA GAAAGTAAAA GAAGAGCTGG
174841 GAAAAGAGAG TGGAGGGGCC AAGGGAATTT TCCCTTTTGG CTTCTGGGGA AACTTTGCTG
174901 AAAAATCAAC TCACAAATTT ATTAACATGT ACACAGGGAG AACCATAGAA TGATTATCCA

```

Figure 8 (Page 54 of 73)

SUBSTITUTE SHEET (RULE 26)

70/162

174961	CTTCCCAAGA	GGGCTTAAAA	GCTTATATAT	TATCCTGGCA	AAACAGATTA	TGGGAGGGGA
175021	AGAAGAGAAA	CTCTGTTGAT	GGGATTACTG	TTGCGGATTT	TTGCTCCTTC	GCTCAGCTAG
175081	GTCCGGGTTT	TTGTCTCACA	GCCAGGAAGA	ATTAGGCATG	CAGCCATCAA	AGAATGAGTG
175141	GAGTAGAATT	TATTAAGTGA	AAGGAAAGCT	CTCAGCAAAG	ACAAGGGTCC	TGAAAGCAGA
175201	TTTCTGGTTT	GCTCTTCACA	GTTGAATACT	AGGGCTTAAG	ACTCAAATTC	CTGACAACTC
175261	CACCTGTGCC	TACCAGTGCA	TGCAGGCCTT	TAGACTGAGC	TACTCCATAT	TGATTAATTT
175321	CCTGAAGTGT	GCATGTGTTA	AGGAAAGGAA	TCATCCACTG	CAGGCATGTT	TAGGCAAGCC
175381	CCCTGTGCAA	GTTCCCTTAT	CTGCACAAAA	CATCCGGTGT	AAGCACTTGT	GGGGCAGGTC
175441	AGAGGTTCTC	TGGGTACCAT	TCCCTTACTG	TCTGCCTAAA	GCAAGCTGGC	CAACTCCTTT
175501	CATTACTAGG	GAGAGTAAGT	AGATCAGGGA	ACAGAGATTA	ACTTGAACAT	TATCTTGTGA
175561	AAGTCCGTTT	GGGCATGGTT	ACATTCTTGG	TCTTACAGGA	AGGGTAAATA	AAAAAATTTG
175621	CTCTTTTGGG	TGGGTCTGGA	TCTTAGGTAG	ATAAAGAAAC	TTTAATTCCA	CGATGTGTTT
175681	TGGTAGGGAT	AGTTGGTGGC	AGGGATGTCA	GAGAGACTTT	GAGGCTTCTT	CAGTTCAATA
175741	TGACCAAGGG	CCATATATTA	GGGTATCAAT	TTCTGAGCCC	CAACAAGAGC	TTAGGAGAGA
175801	TGTGATAGCA	TCACAGTGTG	AAAGCAATTT	TTTGTGTTGT	TTTAGAGACA	GGCTCTTGCA
175861	CTGTCACCTT	GGCTGAAGTA	CAATGGTACG	ATCACAGCTC	ACTGTAATCT	TGAACTGGGT
175921	TCAAATGATC	CTCCCATCTA	AGCATTTCOA	AGTGTGTTGG	TTACAGGCAT	GAGCCACGGT
175981	ACCCAGCCTG	AAACTGCACC	CACTTTCTGA	TAAACTTTTC	AAATGACTAA	AGGGGAGAGA
176041	GTAAGCACTA	CTCAGAGGTA	GGAAGAAAGG	ACACAGGATT	ATAGGATTAA	AACAACAACC
176101	ACCAAAAAAA	ACCAGACCGG	TGTGGTGGCT	CACACCTGTA	ATCACAGCAC	TTGGGGAGGC
176161	TGAGGTGGGG	GGAGTCACTG	GAGGCCAGGA	GTTGAGAGCG	AGCCTGGCCA	ACATAGCAAG
176221	ATGCTGTCTC	TATTAAAAAA	AAAAAATACC	TGCCTTGAGC	TAATCAGAAT	CATGGACCTT
176281	GACAAAGGAT	GTCCCAAAGT	AAGTCTTAGC	ATTTTTTTTT	TTTTTTTGAG	ACAGTCTCGC
176341	TGTGTTGCCC	AGGCTGAAGT	TCAGTGGCGT	GATCTCGGCT	CACTGCAACA	GCTGCCTCCC
176401	AGGCTCAAGC	AATTCTCCCT	GCCTTCAGCC	TCCCAAGTAG	CTGGGATTAC	AGATGCCCCAC
176461	CACCACGCCT	GGCTAATTTT	TGTTTTTTTT	AATAGAGATG	GGGTTTTGCC	ATGTTAACCA
176521	GGCAGGTCTT	GAACTCCTGA	CCTCAAGTGA	TCTGCCCCACC	TTGGCCCCCTC	CATAGTGCTG
176581	GGATTACAGG	CGTGAGTCAC	TGCACCCGGC	AAAGTCTTAG	CATTCTTTAC	AAACAGTTTG
176641	TACCCGTATC	TCTAAAAGGG	AGTAGTGAAT	TTCACCCCAA	AATGTGGCTT	CCTGATATAA
176701	TGAGTATTTT	GAATGAAAAA	CTCTTAGAGA	TCAACAGACA	CTAAAGAGAC	TTTTCCCTAG
176761	GTACATAAAA	ATAGGATGGC	CCCACCAGCG	AGAACAATTG	TTCTTTTCTC	CCTCTCTGTT
176821	ATCTCATTGT	GCATTATAGG	AAAGACCAAG	AATGTAACCA	CACCTGAACA	GACCCTTTTA
176881	TAAGATAATC	AGTCTCTAAG	CATCATTTAA	ATTCCAAGGA	GAATATTATA	CAAAATTTATC
176941	TGTTCTTTGA	TCCAATTAGT	CTCTCCTGGT	AGTTACATAT	TGCCCCCTCAA	CAGAATTCCT
177001	CTTCTTCTGT	TTCCCATAAC	CTATTTTGCA	AGGATCAAGC	CCCTGTTATT	TCTTCAACTT
177061	CAAGGTGGCA	TATAAGCTTC	TAAATTCAC	TGGGATATTG	GTACTATGTG	CATGAGGAGA
177121	ACCACAGAGT	AATTAAATTG	TAAAGCCTTT	TATCTTATGA	ATCTGCCTTT	TTTTGTGTTT
177181	ATTTTTTCAGC	AAAACCTCCA	AGGGCAAAGG	TATAAAACAA	AAATAAAATT	CTAAAGCCCC
177241	CCAACCATCT	GAATAGACTT	TCTCTTCAGT	CAGGCTTCTT	AAAATGTAAC	CTGAAAGACT
177301	GGCTCAGGCC	ATTAAGGGAA	GTGGGGGTG	AACATGCCTC	ATTATTCCTC	TCTGGCATT
177361	ACATCAACAC	AGCTTTTAAG	TCTGATAAGA	AACATTTTAC	AACCTATTCT	CTCTGAAGCC
177421	TGCTAGCTAA	AAACTTCATC	CCATAGTACA	ACTTTGGTCT	TCACAACCTG	TTATCACAAC
177481	CTAGTGCTCC	TTTCTATTAA	TCCCAAATCT	TTATACAAAC	TCAACCAATT	GTCATCACCT
177541	CCACCCCACT	CCTCCGCTGC	TTCCAGTTGT	CCCGCCTCTC	TGGACCAAAC	CAGTGATACAT
177601	TTCTTAAACG	TATTTGATTG	ATGTCCCATG	CCTCCCTAAA	ATGTATAAAG	CCAAGGTGCA
177661	TCCCAACCAC	CTTGAGCGCT	TGTTCTCAGG	ACCTCCTGAG	GGCTGTGTCA	TGGGCCATGG
177721	TCACTCAAAT	TTGGCTCAGA	ATAAATCTCT	TCAAATGTTT	TACAGAGTTT	GGCTCTTGTC
177781	ATGACACAGA	TGACTGCTTC	ACTGAAGCCT	GCTCTGGAAG	TGAGTGGGGG	TTTTGCAAGG
177841	ATAATTTTCC	CCGGATAGCC	CCAGAAGCAG	CTAGTAATAA	TACACTTAAA	GGTAGCTAAA
177901	ATGCATTGAA	CACTTGTTTT	GTGCCAGACC	TATGTCAACA	TTTGCTTTGT	GCCAGGCTTA
177961	TGCCAGTACT	CCTGATTTGT	TAATACATTC	TAAATAAAAA	TTCTGGAGTT	TCAAATATAA
178021	TAACTGAAAA	ACAGAAAATA	AATAAAAATA	TATAATAACT	GAAATAAAAA	TTACTAAGG
178081	CTGGGGATGG	TGGCTCACTC	ACACCTGTAA	TCCTGTTACC	GGAAAGGGGT	CCGTCCAGAT
178141	CCAGACCCCA	AGAGAGGGTT	CTTGGATCTC	ACACAAGAAA	GAATTCGGGC	GAGTCTGTAA

Figure 8 (Page 55 of 73)

71/162

```

178201 AGTGAAAGCA AGTTTATTAA GAAAGTAGAG GAATAAAAGA ACGGCTACTC CATAGGCAGA
178261 GCAGCTCTGA GGGCTGCTGG TCGCTCATTT TTATGGTTAT TTCTTGATTA TGTGCTAAAC
178321 AAGGGGTGGA TAATTCATGC CTCCATTTTT TAGACCATAT AAAGTAACTT CCTGACGTTG
178381 CCATGGCATT CGTAAACTGT CGTGGCGCTG GTATGAGCAT AGCAGTGAGG ACGACCAGAG
178441 GTCACCTCTCA TCGCCATCTT GGATTGGTG GGGAGCAGTG AGGATGACCA GAGGTCACCTC
178501 TCATCGCCAT CTTGGATTTG GTGGGTTTGA GCCAGCTTCT TTACTTTTTT CTTTTTTTTT
178561 TTTGCCCAGG CTGGAGTGCA GTGGCAGCAT CTCAGCTCAC TGAACCTCC AATTTCTGAG
178621 TTCAAGCGAT TCTCGTGCTT CAGCCTCCCA AGTAGCTGGG ATTACAGGCA TGTGCCACCA
178681 CACCCAGCTA ATTTTTTATA TTTTAAATAG AGACCGGGT TCGCCATGTT GCCTACGCTG
178741 ATCTCCAACCT CCTGCGCTCA AGCCATCCAG CCACCTTAGC CTCCCAAAGT GCTGGGCTTA
178801 TAGGTGTGAG CCACCCACC TGGCCTAGCC GGCTTCTTTA CTGCAACCTG TTTTATCAGC
178861 AAGGTCTTTA TGACCTGTAT TTTGTGCCCA CTGCCTGCCT CATCTGTGG CTTACAATGC
178921 CTAACCTTACA GGAATGCAG CCCAGCAGGA CTCAGCCTTA TTTACCCAG CTCCTATTCA
178981 AGATGGAGTC TTTCTTGTTT AAATACCTCT GACAAGCCCA AACTTTGGG AGGATGACAC
179041 AGGAGGATTG CTTTAGCCTA GGAGCTCAAG ACCAGCCTGG GCAACACAGT GAGACCCCAT
179101 CTCTAAAAAA AAAAATACAA AAAAATTAGC CAGGCATGAT GGTGTGTGCC TGTAGTCCCT
179161 GCTACTCAGG AGGCTGAAGT GGAAGATGG CTTACGCCCA GGAATTCAAG GCTGCATTGT
179221 CAGAGGCATT TGAACCAGAA TGACTCTATC TTGAATAGGC GCTGGATAAA ATAAGGCTGA
179281 CACCTGCTAG GCTGCATTTT CAGTATGTTA GGCATTCTTA GTCACAGGAT GAGATAGGAA
179341 GTCAGCACAA GGTACACATC ACAAAGACCT TGCTGATAAA ATAGGTTGTG GTAAAGAAAT
179401 TGGCCAAAC CCATCAAAAC CAACATGGCC ACCAAAGGGA CCTCTGGTTG TCTTCACTGC
179461 TCATTATATG TTAATTATAA TGTATTAACA TGCTAAAAGA CACTCCTACC AGCATCATGA
179521 CAGCTTACAA ATACTGCGGC AATATCTGGA CTTTACCTTA TATGGTCTAA AAGGTGGAGG
179581 AACCCTCAAT TTTGGGAATT GTCCACCCCT TTTTGGGAAT GCTCATGAAT AATCCACCCC
179641 TTGTTTAGCA CATAATCCAG AAATAACTAT AAGTATGCTT ATTTGAGCAG ACCACGCTGC
179701 TGTTCTGCCT ACAGAGTAGC CATTTCTTTA TTCTCTTACT TTCTTAATAA ACCTGCTTTC
179761 ACTTTACTGT ATGGACTTGC CCTAAATCTT TTCTTGTGTG AGATCCAAGA ACCCTCTCTT
179821 GGGGTCTGGA TCAAGACCCC TTTCTGGTAA CATCTTCTG GTGACCACGA AGGGACAATA
179881 CTGAGGAGAC TCTGAAGCCA AAGGAAACAG ACTACAGCAC CAACTGGCTG ACTTTGGGTA
179941 AGTGGTGGAG TCCCCGGGTA AAGGATAGGA TTGGGTTAGA GGTGCAACTT AGGGGAGATA
180001 GGGTCTCTCC TAAGACAGAG AGGGTTTCAG TCCGCTCTTA ATAAAGGGCA AGAATGCTTG
180061 ACCGAACCTG GGTTTGAGAC CCAACTTAGG AAGGCTACAG TCCTTAAGAT TTAAGGGGTT
180121 AGAGGCCCCCT CTCAGTAAAG TCTCTCTTGG TTAATAACGG ATTTAGCATT AGGGGATGTT
180181 AACTGCTATT CTGTTTGAT TAATCTTCCC TGTGCTCTTT GCTGACAGCT ATGGGTGACA
180241 GGATTAGGCA TGTACAGGAT CACGGGACAT TGGGAACCTT TCTTCTCTCC AAAAGGGGAA
180301 CTTTGACAGC TGATAGGACT GTTGGAAGAG ATCCCTTTGC TATGACAAGC AGCCGCTGA
180361 ACTTTTGATT CAGTGTGCT GCAATGGGTG GGTCTTCTC TGGCCTCTGT GAACCTCTCA
180421 CCTTCCCAT CTCACCACAG GCAATGCTTT TCTCCCTTTC TCTCTTTTCT CTTTTCTGTC
180481 TTTTCTGTTA CTTGAGACAA CCATCTTGCC CAGAGACCAT ATGTTGAAAC TCCTGGTCAG
180541 AAGTTTGATT AAAGATGAAA GGGCTATCT GGGGGCAAGT TTGAGCCTTC CCAGTTAGAT
180601 ATTGGGTGCT AAGTGGAGTG GCCAATGTCT ATGTTTTGTC ACATGTATAT TGCTCTGGCT
180661 GAAATGAAA ACGTTAATTT GGTACTTTA TGTGGCCATT GGGCAGCATC TTACAAAAGT
180721 GAGAGACATT TATTTGCCTG TGGTTCCATG AAACAGAAA AAGTTGGTTT TCTTTTGTGT
180781 CGTAGCTTGG ACCCAAGGGC TTTGCAGTGA GCAAGGTGTC TAGTGCTGCT CAGTGAAAGA
180841 GAACCCAGAA ACCTGGCATG CCAGCAAAAG GGTAAAGATT TCTTACCAGT CAGGCTTCTG
180901 GCCTCTCTCT CTTAGTGAAA ACTGAATGAA TGGTAAAAAT CACTGTTTAT CACCTCTGTA
180961 AAGTTTTGAT TAATGGGAAC AAGGATTTGT GGGGCTAGTC TTAAGCTGTA ATGAATCTGG
181021 TATACTTTGT GATATCAATT TGTCTTCTG TATTACTCTG TCATAAGAG GAATATGGTA
181081 GGATAGAACA TGGGCTCAGG ACTCCATAAG CCTGCTGTTT AAGCCAGCCC AGTAACTGG
181141 TCCGTTGCAA AGTTTATTAC AGGTCCCTGG AAAAAAAAAA AAATAAAAAC TGGATGAAGT
181201 TTCCTTCTCA TCTTGTTTTA TGCTCTTGG AGCTTCACCT TGTAACCACG TGGCGGTACT
181261 TTCTCTTGGT CTCTGCCATC CAGGGAACAG GAATTTTGGG GTTTATGTAA TAGTTAACTC
181321 TAAATAATTAT CTCAAGCCAT TGCAAGCTCA AAATTGGCTG CTCTGGACCC CTTCTGGGAA
181381 GGGCAATGGA AACTAACCAG TGTTGTAGCT CAGCAGCTAA GGATTTGTCA TTTTATAATG

```

Figure 8 (Page 56 of 73)

SUBSTITUTE SHEET (RULE 26)

72/162

```

181441 GCGGCCAAGG TTCAATCCTG GCTTAGGGAA TGAGTACTTT CTGATTGATA TCTGTGTGAC
181501 CTTTACCATT TGTTGATTCT GTTCTCTTCC CCTCCACACA CTGTCTTGAG TTTTCCTCTC
181561 TCTGAGAACC TGGGAGATTA TCTTTGGTAA AGTTCAAAAG CCAGAAATAA TGGCCGTGTG
181621 GGATGGCTAA AGTTGAGTAA TAAGAACTT AAAAGGACTC CTTTTTTTTT TGCTTTAGAG
181681 TGCTATGGTT TATGGTTAAA AGCTTAATTA AAAGTGGATA TTCAATCTCT AAAAGCCTGG
181741 GACTCCTTGG GAAAAGCAGA GGAGGCACCA CAGACCCCAT TTTGGGAAAA CCTCTGTTTT
181801 CCTCATGAAA CCCCAGGAAC TGAAGTGA TAGATCCTTC GCAAAATCTA AGGCTCTGTT
181861 TGGCTTTGCA TTATGTTATC TGATGTTTTT GACTTTTGGG GGTATCAGAA ATTACTTTGC
181921 ATTATGAGGG AGATCTGGTG TGTAAATACC AGGTAGGAAA TATACTTCTG GGGATAGCTA
181981 AAGGCAATA TAGGTGAATA CTTGGCTATT TGCACTTTGG GATCACAAGA AGCATCTCT
182041 TGACTACCTA GAAGGTATGG AAATGTCTCC ATCCCCACCG AGAGATAAGA TTCCCAGGGG
182101 AGATGGCTGA TCCCCAAAA GAGGGCTGAT TCCCTCTTTT GGGATCCAGG ATCTGGTATA
182161 AAAATGGGAC CCTGGCCAGG CACAGTGGCT CACGCCTGTA ATCTCAACAC TTTGGGAAGC
182221 CTCAGAGTTA TGAATGTCTC ACCATACTGA CACTTTGTGA CTGAGCTCCT CTCTACCCTG
182281 GACACAAGAG ACCCTAATAA TTAGACAGGA ATATCATTGC CCCTATTAG TCTGAAGAAG
182341 TTATAGAAGA CGGATCTTTA TCCCCTGCA ATCCTTAGGA TTAAGGGTTC CCTGGTAAAA
182401 GGGAGTGGGA AAATATGTCA GAGGCATTG AATCAGAGTG ACTCCATCTT GAATAGGGGC
182461 TGGGTAAAAA AAGGCTGAGG CCTGCTGGGT TAGGTAGGC ATTCTAACCA GGAGTTTAGT
182521 CACAGGATGA GATAGAAGGT TGCACAAGGT ACCCGTCACA AAGACCTTGC TGATAAAATA
182581 GGTAACGGTA AAGAAGCCAG CTAAAGCCCA CCAAAACCAA CATGGCCACA AAAGTGACCT
182641 CTTGTCATCC TCACTGCTCA TATACACTAA TTATACTGCA TTAGCATGCT ACAAGACACT
182701 CCCACCAGTG CCACGACAGT TTACAAATAC CATGACAACA TCTGGACGTT ACCTTATATG
182761 GTCTAAAACG GGGAAAGAACC CTTAGTTCTG GGAATGTCC ACCTCTTCC TGAAAAATTC
182821 TTGAATAATC CATTAGTTTA GCACATAATC CAGAAATAAC TATACGCTCG CTTATTTGAG
182881 CAGTCCATAC TGCTGCTCTG CCTATGGAGT AGCCATTCTT TTCTTTTATT TTTATTTTTT
182941 AGATAAAGAC TCGCTCTGTC ACTCAGGCTG GAGTCTGGAG TGCAGTGACG TGTTTTGGCT
183001 CACTGCAACC TTCACCTCCC GGTTCAAGC AATTCTCCTG CCTCAGCCTC CCAACTAGCT
183061 GGGACCACAG GTGGGTGCCA CCATGCCCTG CTAATTTTGG TATTATTAGT AGAGATGGGG
183121 TTTCGCCATG TTGGCCAGGC TGCTCTCGAA CTCCTGGCCT CAAGCGATCC ACTTGCCTTG
183181 GCCTCCCAA GTGCTAGGAT TACAGGCATT ACCCACTATG CATGACCCAT TCTTTTATT
183241 CTTAACTTTT TTTGTTTTTT TTGAGACAGA GTCTCACTCT GTCACCCAGG CTAGAGGCTG
183301 GAGTGCAGTG GTGCGATCTT GGTTCACTGC AACCTCTGCC TCCTGGGTTT AAGCGATTCT
183361 TCTGCCTCAG TCTCCTGAGG AGCTGGGACT ACAGACATGT GCCACTACAC CCAGCTAATT
183421 TTGTATTTTT AGTAGAGACA GTGCTTGCC ATGTTTGTC GGTGTGCTC GAACCTCTAA
183481 CCTCAAGTGG TCTGCCTGCC TCAGCCTCCC AAAGTGCTGT GATTACAGGC ATAAATCACT
183541 GCGGTCGGCC CTTCTTTACT TTCTTAATAA ACTTGTTTT ACTTTACTGT ATGGACTAGC
183601 CCCAAATTCC TTCTGTGTG AGATCCAATA ACCCTTTTGT GTGTGAAAGA ATGTATTGCT
183661 GCTGTTTCAG CTGGAGCAAG CTGGAGCTCA TGCTGCTGCT CAGACTGGAG CATGCGTGAT
183721 CTGTGATCCC AGTAAGAGGA TCATGGTCAC TCCAGCCTGA ACGACAGCAT GATATCTCAT
183781 CTGTAAGAAA AAAAAATTAC TAGAGGGCTT TAACAGCAA TTTGAGCAGC AAAAAAGAAGT
183841 AATCAGTGAA CTCAAAGATA GGTCAATTGA AATGATCTAC TCTGAAAAAC AGAAAGAAGA
183901 CAGAATGAAG AAAAAGAAAT AGAGCCTTAG AGACAGGGGA TACCATCAAG CATACTAATA
183961 TATGCATAAT GGGACTCCTA GAAGGAGAAA AGTGAGAGGA CAGGGAGAGA GAATGTTTGG
184021 AGAAATAATT TCTCAAAGCT TCCCATGTTT GGCAAAAAAG CATTAAGTTG CATACATATT
184081 TTAGGAGCTC AATGAATTCC AAGTAGGATA CACTCAAAGA GATCCATACC TAGACACATC
184141 ATAATCAGAT TATCAAAGA TGAAGAAGAT GAATCTTGAG AGCAGAAAGA AAGGAACAAT
184201 TCATCACATA CAAATAGTAC TCAAAGATG TCTGGAGTAG GTATACTAAT ATCAGACAAA
184261 ATAAACTTTA AGATAAGCAT TGTTATAATA AATAAAGAAA GGTATTTTGT AATGATAAAA
184321 GTGTCAATTC ATCAAGAAAA CATAACATTA TAAACATACA TGCACCTAAC AACAGAGCCC
184381 TAATATTCAT GAAACAAAAC TGACAGAATT GAAGGGAGAA ATAGAAAAAT CGACAATAAT
184441 AGTTGGAGAC ATCAATACCT CACTAGTTAG ACAAGATCAA CAAAAAATA GAAGACTTAA
184501 CACTTGAAAA CACCTAACCT GACCCTAACA TAAATCTATA GGTCACCTACA CCCCCAACA
184561 GCAGAATAAA CATCCTTCTG AAGCTCATAT GAAACATTTT TCAGGATAGA CTGTATATTA
184621 CTTTCATGAAA TAAGTCTCAA TAAATGTATA AGGACTATAA TAATAGAGTA TATATTCTCT

```

Figure 8 (Page 57 of 73)

SUBSTITUTE SHEET (RULE 26)

73/162

184681 GACCAAAGTG GAATGAAGAT AGAAATCAAT AACTAGGCTG GCGTGATGG CTCACGCCTG
184741 TAATCCCAGC ACTTTGGGAG GCCAAGGCGG ACAGATCACG AGGTCAGGAG TTTGAGACCA
184801 GCCTGACCAA CATGGTGAAA CCTGTCTCT ACTAACAAA TACAAAAATT AGCCAGGCCT
184861 GGTGGCATCT GCCTGTAGTC CCAGCTACTC GGGACACTGA GGCAGGAGAA TCACTTGAAC
184921 CCAGGAGGCA GAGATTGCAG TGAGCTGAGA TCGCGCCACT GCATTCCAGC CTGGGAGACA
184981 GAGCGAGACT CCGTCTCAAA ATTAAAAAAA AAAAAGAAAC TAGAAAAATA AGAACAAATC
185041 AAACCCAAAG CAAGCAAGAG AATTTCAAG AATTTCAAG CAGCCAAGAA CAAAAGGCAC
185101 ATTATGTACA GAAGAACAAG TGTATAGATC ACATATTTCT CATAGACACA ATATAAGCAA
185161 AAAGACAGTG GAGCAAAATT TTTTAGATTA ATGAAAGACC TACAATTCTG TACCAAGCAA
185221 AAAAATCCC CCCAAATGAG GGTGAAATAA GACAATTTAA TACAGAGAAA AGAGGAAGGA
185281 ATTTATCTAG TCATATGTGA GAGTTTATG ATACATTTTG TACTGTATAT GTGGATGTTT
185341 TCTATTTTCAT TTAAAAATC AACCCTGCAA TTAAATGGTA GATTGTCTTG CTTCTTTTGT
185401 ATTGACACAG TCATTAACCTA AAATATTGTA GTATTTTTTT ATCTCCCTGC CTAAAGGCAA
185461 TAAACATCTA ATCAGCAGAC TAGAACAATA AAAAATATTT TTTAAAAGTC CTTTAGGCAG
185521 AATGATAAAA GTCCCTTAGG CATATTGAAA TTCCTATTTA TACAAAGGAA TAAACAGTAC
185581 TAGAAATTGT AACTATGTGA GTAAACAGAT AATATTTTTT CTCCATAAAA TGTGGTTGAC
185641 TATTTTTCACA AAAATAGTTA ACAATGTAAT GTGTGATTTA TAGCATTTAA AAGTAAAAACA
185701 GGCCGGGCAC AAAGGTTTCGT GCCTGTAATC CCAGCACTTT TGGAGGCCGA GCGGTGCAGA
185761 TCACTTGAGG ACAGGAGTTC AAGACCAGCC TGGCTAACAT GGCAAAACCC CATCTCTACT
185821 AAAAATACAA AAATTAACCA GCGTGGTGG TGCACGCCTG TAATCCCAGC TACTCTGGAG
185881 GCTGAGGCAC AAGAATCACT TGAATCCAGG AGGTGGAAGT TGCAGTGAGG CAAAATTATA
185941 CCACTGTGCT CCAGCCTAGG CAACAGAGCT AGACTCTGTC ACACACACAC ACACACACAA
186001 AAGAAAAGTG TATGACAACA ACAGTGCAA AGAAGTGGAA ATGAAAATAA TGTTATTTTA
186061 TATAAGTGGT ATACTTTTAG ATGAATACG ATAAATTAAT GATGTATCT ATAACTCTA
186121 AGGCAACCAC TGAATAATG AAACGAAGAA TTATGGCTAA CAAGCCACAA AAAGAAATAA
186181 AATAGAATGA GAAAAAATAT TTAAGTTGTT CAACAGATGG GAAAAAAAAG AGGAAAAAGA
186241 GAACAAAGAA CAGATGGGAC AAATGGGAAA GTAATAGCAA GATGATAGAC TTAATCTTAC
186301 CCATATAGAT TATCACACTT AAGGTAAATG ATCTAAATAC TCTAATACAA AAGCAGAGGT
186361 TGTCAGATTG AATTAAAAAA ACAGACAACA AAAAAAAAAG GCAAAAAAAG AGCCACAACA
186421 TGCTGCCTAC AAAAAATTCA CTTTAAATATA AAGACACAAA TAGCTAGAAA CACCATCACT
186481 TTTAACCTTA TTTACTCAAA CCTCCTGATC CCTATTTATT TATTTATTTA TTTATTTATT
186541 TATTTATTTA TTTATTTATT TTTGAGACAG AGTCTGACTC TGTGCCCAG GCTGGAGTGC
186601 AGTGGCACC TCTAGGCTCA CTGCAGCCTC TACCTCTCGG GTTCAAGCGA TTCTCCTGCC
186661 TCAGGCCTCC CAAGTAGCTG GGACTATAGG CACATGCCAC CATGCCAGC TAATTATTAT
186721 ATTTTATAGTA GAGACGGGGT TTTGCCATGT TGGCCAGGTT GGTCTCAAAC GCCTGACCTC
186781 AGCCTCCCAA AGTGCTGGGA TTACAGGCGT GAGCCACAGC ACCCAGCTCC TCTTCATTTA
186841 TTCTTGCTAC GCTTCCTCCA ATCCATTTTG TGCATTTGAT GATTTTGCCA GTAACCTCTT
186901 TATTTTCTG GTAAAAATAC TTATGGGTCA CTGAGGACTG GGATGTTCTT TCTTCTAGAG
186961 GGGGTTTGTG TCTGCTTTG CCAGGAAGCT GGGGTACCAC CAGTCAAGTA TTACTTTAAA
187021 CTCAATTCAT GAATTGAGAC TTTTTTTTTT TTTTTTTTTT TTACGCAGAG TCCTACTCTG
187081 TCACCCAGGC TGGAGTGCAG CCGTGTGAAC ATGGCTCACT GCAGCCTCAA CCTACTGAGC
187141 TCAAGCAATC CTTCTGCCTC ACCATTCTGT ATAGCTAGGA CTACAGGTGT GTGCCACCAT
187201 GCCTGACTAA TTTTTTAAAT ATTTTTTTTA GAGATGGGGC TCACTTTGTT GCCCAGGCCA
187261 GTCTCGAGCT CCTGGGCTCA AGTGATCCTC CCACCTTGGT CTCCCAAAGT GCTGGGGTTA
187321 CAGGCATGAG CCTCTGTGGC TAGCCAAGAC TTTTATTTT TTAGCCTAAA TGTGTATAAA
187381 AGTTGGCTTG TGGTTACAAC TTATCAGGAT TGATGATCTC TCTCTCTCTC TCTCTCTCTC
187441 TCTGTCTCTC CCCACCTCTC TCACATCCCT TGCTCTGCTG AGAAGCAGAG CAAACATTCT
187501 AGCAGTTTCC AGAGAGTAGG ATGGGATTAC TTCTAGTTTA CTTTTATCAT CTTTGGGAT
187561 CGCAGTATTA CTGGGAGAAC ACAAGTATCT CTTATTAGAC ATACCACCTT TGTAGAATCT
187621 GGACTTTTAT TTTAGACTTT ATTTGTTTTT TACTATAAGC AATTTAAGTT ACAGATCTCT
187681 CTACACACTG TTTAAGTTGC ATCCCATGAA TTTTGATGTG CTTTATGTG ATTATTATAT
187741 AGTACAATGT ATTTTGTAAAT TTTTGTGAT TTGTTGGAG AGATTGATTA ATTAGAATGA
187801 TGTTTAATTT CCAATATATG GTGTTTTTTT CTACATTTCT TATTTTATT GATTTCAAAT
187861 TTATTTCTAC TGTAGTCAGA TTTAATAATT CATTTATTTT TATTATTTTC ATTTTTTTAG

Figure 8 (Page 58 of 73)

74/162

```

187921 AGACAGGGCC TTTCTGTGTT GCCCAGGTTT GTCCCAAACCT CCTAGTCCCA AGCAGTTCTC
187981 CTGCCTCAGC CACCCAAAGT GCTGGGATTA TAGGCACGAG CCACCCGTGC ACAACCAACA
188041 ATTCAATTAA AAAGTGGGCA AGTGAACCTGA ACAGACATTT CTCAAAAGAA GGCATACAAT
188101 TGGCCAACAA ATATATGAAA GAATGCTCAA CATCACTGTA TTAGTCTGTT TTCATGCTGC
188161 TAATAAAGAC TTAACCTGAG ACTGGGGAAT TTACAAGAGA AAGAGGTTTA ATGGACTTAC
188221 AGTTCCACAT GGCTGGAGAG ATCTCACAAT CATGGTGGAA GGCAAGGAGG AGCAAGCTAC
188281 ATCTTACATG GATGGCAGCA GGCAAAGAGA GAGCTTGTGC AGGGAAACTC CCGTTTTTAA
188341 AACCATCAGA TCTCGTGAGA CTCATTCACT ATCATAAGAA CAGCATAGGA AAGACCCGGC
188401 CCATAATTCA GTCACCTCCC ACTGGGTTCC TCCCAGGACA CATGGGAATT GTGGGAGTTA
188461 CAATTCAAGA TGAGATTTGG GTAGGGACAC AGCCAAACCA TATAAATAAC TAATCATCAG
188521 GGAAATGCAA ATCAAACCA CAATAAGGTA TCATCTCACC CCAGTTAGAA TGGCTATTGT
188581 CAAAAAACA AAAAAATAACA AATGCTGGTG AGGATGTACA GAAGAGGGGA CTCTTATGTC
188641 CCACTGGTGG AAATGTCAAT TAGCATAGCC ATTATGCAA ATAGTATGGA AGTGAGGTAG
188701 GTTACATAGG GTGGTCACAG CCTCCCTTGA AAGGAAACAA GAACTTGTC AAATTGATGG
188761 AGAGAACAAA TCTCTTGACA TTACACAAAC TGCATCTGGG GCTAGTGGTT AGAATATCCT
188821 CAGTCAAGGA GGTAAGAGAG CAGGAGGGAA AATCCCTAAG TTCGTGCAAG TGCAGAAACC
188881 CACAAGCTGT GTTCTCAGGT TGACATATAC TCATTTTAAT AGTAAGAAAC ACACCCTTGG
188941 GTAGAGAATT AAAATGCTAA TAATACATGT GATGTATGTA CTAGCGTGTA TGGCAATATT
189001 GCATGCACAT TCAAGAGACC ACCCAAACA TATTTAACA CAATGCCCAT TCCCACCCCC
189061 TCATGGATAA TCACGTAGGA CTCCATAAC GGGAGTTTCT TCAGTGTCAA TTGGTGCTGA
189121 AGTAGCCGAC CCTGACTCTG CTATCAGCGT GTACTTTCAC CTTGCAATAA ACTCCTTTGC
189181 CTACTTTTAC TTTGGACTGG CTTTCAAATT CTTTTGTGCA GGGAAATCAA GAATCTGAAC
189241 CAGCCTACTG ACAACAGAGG TTTCTCAGAA ACCTAAAAAT AGATCTACCA GATGAGGCTG
189301 AAAATCTGCT ACTGGCTATT TATCCAAAGG GAAGGAAATC AGTATACAAA GAGACACCTA
189361 CATCCCCATG TTTATTGCGT CACTCTTCAC AAGAGCTGAT ATATAGAGTC AACCCTAAAT
189421 GTTCATTAAC AGACAAATGG ATAGAAAATG TGGCATATAT ACACAATGAA ATACTATTTG
189481 GCCATGAGAA GAATGCAATC TTGTCATTTG TGGCAACGTA GATGAACTG GAGAACATTA
189541 TGTTAAGTAA GATAAGCTAG GATTGGAAAG ATAAATACTA CATGTTATCA CTCATATGTG
189601 AAAGTAGAGA AAAATTTTTA GCTCATGGAT TTAGAGAACA GAACTGTGGG TACCGGAAGC
189661 TGGGAAGGGT AGCAAGGAGG GGAGGATAGG GAGAGGTTGG TTAATGGTGA CAAAATTACA
189721 GCTAGATTGT AGAAATGAGT TCCGGTGTTT TGCACCATTG TAGGGTGCAT ATGGTTAACT
189781 CTCATTTATT GTATATTTTC AAAAGCTATG AAAAGAATTT TGAATACTCA CAACAAAATA
189841 AATGATAAAT GTTTAAGGTG ATGGATATAC TAATTACTCT GATTGATTA TTACACATTG
189901 TGTACACATA TAAAAATATC ACTCTTTATC CCGTATATAT GTACAGTTAT TATATGTCAA
189961 CTAAAAATAA AAGAAAAAAA GAATATGATC TATCATGATG TATATATCAT GTGTACTTGA
190021 GCAAAATGTG CATGCAGATA TTGTGTATAA TGTTCTATAA ATCAATTAGC TCAAGATAAT
190081 AGATAGGATT GTTCAGATCT TCTGTGTCTT TACTGATATT TTGTCTAGTT ATTGCATCAT
190141 TACCAAAAAA AGGGTGTTAA ACTCTCCAAA TGTGATTGTA GAATTGTCTA TTTTGTCTTT
190201 TCTTTTCCAT TTTTACTTTA TGTATTTTGA AACTCTGTTA TGACATTTTG CTATGTATTT
190261 TAAAACTTCG TTATGTATTT TGAAACTCTG TTGTTAGAAT CATACATTTA TGATTATTAT
190321 GTTTTCTTGA TGAAATGACA CTTTCTATT GTCAATTGTT TTGTTTTTTC TGAATGGAG
190381 TCTCACTCTG TTGCCCAGGC TGGAGTACAG TGGCACAATC TTGGTTCAC GCAACCTCCA
190441 CCTCCTGGGT TCAAGCGAGT CTCCTGACTC AGCCTCCAAG TAGCTGGGAT TACAGGCATG
190501 TGCCAGCATG CCAAACCTAAT TTTGTATTTT TATTAGAGAC AGAGTTTCAC CACGTTGGCC
190561 AGGCTGGTCT CGAACCTCTG ACCTCAGGTG ATCCGCCAC CTCGGCATTT TTATTTTATT
190621 TTATTTTTTT GAGACAGAGT CTCACTCTGT CACCCAGGGT AGAATGCGGT GGTGTGATCT
190681 TGGCTCACTG CAACCTCCGC CTCCTGGGTG CAAGCAATC CCATGCCTCA GCCTCCCGAG
190741 TAGCTGGGAT TACAGGCACA TACCACCATG ACTGGCTAAT TTTTGTATTT TTAGTAGAGA
190801 TGGGGTTTTT CTATGTTGGC CAGGCTGGCA ACTGACTCCT TTAACAATAC AAAATATCAC
190861 TCTGTCTCTG GTAACACTCT CTGTCTTAAA CTCTATTTTA GCTGTTATTA TTATGCCAT
190921 TTTAGTCTTT TTATGCTTTC TGTTTGCATA GTGTATATAT TTTAATATGT TTATTCTCAA
190981 GTTATCTGTG TTTTATATAT TAAGATGTTT CTCTTCTAGC CAACGTGTTT GGTCTTGC
191041 TTTTTAAGTC GATTCTAACA ATCTTTGCCT TTCAATTGAA ATATTTACAC CATTAACATC
191101 TAACATTAAC ATTTATTTTT CTTTCCACAG TACACTGGCT AGCATCTCCC ATATAATATT

```

Figure 8 (Page 59 of 73)

SUBSTITUTE SHEET (RULE 26)

75/162

```

191161 GAACATAAAG TGTGATAACT GACATCCTTA TTTCATTCTT ACTCTGAGTG GAAAGGGCAG
191221 GGGTGGAGAA AGCATTCAAC AATTTGCCAT AATTATAATG CTTTTTGTTA CACTGTTTTT
191281 TTCTGCATTA AAAAATATCA TTACATTTTG CATGAATTAT TAGGAGAAAA TATTTTCCAA
191341 TTTTCTGGA AAATGCCATA ACCACGTCTC TCAATTTTGT TTCCATCTTT CTTCCACATT
191401 TTACATAACC TACATAAGAG ACACATTATC AAGTATATTT TACATGGCTT CTCAGTGTCT
191461 TCTCTGTCTG CTAACAGGTT TACCAAGAGA TGGCACTCTT GTATTTCTGG TGGCTATGTC
191521 CATATCGTTT TGCCTTTAAG ACAGCGTAAC TACTTCTTTC ACCAGTATTA AAGACATGTA
191581 CATTTGATCT GGTTCCTGTG GATGATTTTA AATGACTCAA GCTAATAATC CTAATTTTAC
191641 CTAAACACTC CATTATTTTA AAATGTATTC CTTTATGCCC ACAATAAACA TTTATTGACA
191701 TTAGGCTGGA CATTAGGCTT CTCTATGGCA GACATTAGGC TGGACCCTAG CCATATATCT
191761 ATTGAGGGAA AAAAATTAT TTTCTATATA AGTTTCCAGA AAGCCAAGAT GTGTTTTTAA
191821 AACAAACAA AACATTACAT TCTAAATGCT GTAACAAGAT AAGAAAAAGT GTTGAGGCTG
191881 AGAGAAGAAC AAAGCAGCAA GCAACTCCTG GAAGGACCAC TGCTGCAGAG GTAATAACTG
191941 GTGAACCATG TTTTGGAGAA GGAAAAGGTC ACCAAGAGAA GGAGGGGGTC CAGGGTGTTC
192001 AGAAAGATTG CATGCATAAA GATCAAGGGT AATAAAAAAA ATTCCGTATT ATGTAATGT
192061 GAAGTTCAG GACCATGAGC TTGGAGAGCA TGAAGTACAG GAGGAGGGT GGTTTTCAAT
192121 AAATCTGGGA ATGAACAGT GAAGCCTCTG GCAGAACTCA CATCTCTTTC CTCCCCTCTT
192181 CCTTGACAT TCCCTTTATG GAGTAATTGC AGGGATGGGA AAAGTTCAAA ACCACCACTG
192241 AGCCTAGGAA GTGCTAGGGT AAAGTGGAGA ATGAACCTGC GTGATTTGCT CATCCTAAAC
192301 TAGGTTCTTC TAGGAGAGCC CTTCCCCATA AAATCTGCCC TCCTCGAAGG GGCCAGACA
192361 GCCTAAGCTC ACCTCCCAA GACCCCTTAC TTGCTGACTG AATCTGATTC CACCCAGACA
192421 TGGCCTAAA CCCTTCCATA ACTCTATAGC CAAATTCAAT TTTAGACAGG CCTCATACCA
192481 ACCTTTCTTC CTCTAAGTCT GCCACCTTAG GCAATTCTCA ACATTCTCTA CACACTTTGG
192541 GGCCATAGAC GTGCTACCAA GTCTCCAGAC CTAGACCTGA TGGAGCAGTG CTGTAATGAG
192601 ACGACCACTG GCCTTTGAAC CAGACCCTTC TCTGTGGCTC CTATGCATCT CCAACCTGTT
192661 TTGAGCACTG CTGCCAAGAC ATCTTTTGGA CTTTGTGTG AAGTTTTTAA ACTGAACATA
192721 TCTACAAAAC ACCTAACCTT TAAAAATTCA TTGTCAATTC ATATCATGAA AGATAAAGAA
192781 AGGCCAGGAA ACTGTTCCAG GTTAATAGAG ACTAAAGAGA TAGCAACCAA ATGCAATTTG
192841 TGATCCTGGA TTGAGGGGAA AAAGTGTGT CAGAGACATG ATTGGGACAG CTGGTAAAT
192901 TTGAATTTGA ATTTAAAGAT AAAGTATTGA GTAATATAGG AAGATGATTA TCTGCAACTT
192961 TCAAATGTTT CAGTAAGTAT ATATATATAT AAAGAGATAT AAAGACATAT AAATAAATGG
193021 ATAGGTAGAG AAAAAGCAAA TGTATAATAT TAACAATCTA GGTAAAAGT ATATGAGTGT
193081 TCTTTGTACT GTTTTTCTGA TTTTCTATA TGTTTGAAT CATTTTAAAA TAAGAAGGTT
193141 TTTGGGTTT TTTTGTGTG TTTTGTGTT TAGAGACAGC ATCTTATCTG GTCACCAGGC
193201 TGTAGCTCAG TGGCCCAATC ATTGCTCACT GCAGCCTCAA CTTCTGGGCT TCCAGTAAT
193261 CCCCCTACCT CAGGCTCATG AGTAGCTGGT ACTTCAGGTG TGCACCCTG CACTCAGCTA
193321 ATTTTTATTT TTTAAATTTT TGTAAGATG GCATGTTGCT ATGTCACCCA GGCTAGTCTC
193381 AAACCTCTGC CCCCAGTGA TCCTCCCACT TTGGCCTCCC AAAGTGCTAG AATTATAGGC
193441 ATGAGCCACT GCACCCAGCC CCAAATAAAA AAGTATTTTA TTTTAATTAA CTAATTAACT
193501 TTGAGTCAGA GTTTCACCCT TGTACCCAG GCTGGAGTGC AATGGCATGA TGTTGGCTCA
193561 CTGCAAATC TGCTCCTGT GTTTAAGCGA TTCTCTTGCC TCAGACTCCT GAGTAGCTGA
193621 GATTACAGGT GCCTGCCACC ATGCCAGCT AATTTTTATA TTTTATAGT AGACGGGGTT
193681 TCAGCATGTT GGTCAAGCTT GTCTCAAAT CCTGACCTCA GGTGATCCAC CCACCTCCGC
193741 CTCCGAAAGT GTTGATGAGC CACCACACC GGTCTAAAAA GTATTTTAAA ACCACAGTCC
193801 CACTCTACCT TGTCTACAC TACCAGGGGC TAGGATCACC CCATGTCTTC TAGGCTATGA
193861 GATAGAGGAA TCCAAGGAAG AAGATAAGCT ACTTGGTTCC TCTATAGGGT CTTGTGTGTG
193921 CTCTCATGTG CTCTCTCTCT CTCTCTCTCT CTCACACACA CACACACACA CACACACACA
193981 CACATGAATA CCAGAGCTAT CACTTTCCCA GTCTAGTACT CATCTCATCC CAAGGGTTTT
194041 GTGTTGTAGT GGTGTGCTCA TTTCTTTGTT TTGTTTGTGTT GCTTGGATTA TTCTTTTTCT
194101 CTTTTTGCAG CTGAAGGGAG AATTTCCAGG CCAGCCCTTT GGCCATTAGA GTTACAGTGC
194161 CTCTATTGAG GCTTCATAGA GAGACCTGGG ATTCAGTAGT GGGGGGCTTT TATCCAGTTC
194221 AAAATAATGC ATTCTACCA AGATGTACTT TGAAATAAAA CAATACTAAA ACACAAAATT
194281 TTATTTATGC TGAACATTGA ATCACTTTTT TCTGTATTTT GTGTAGAAAG TTATACACAC
194341 ACAAACACAT TTGCTCCTGC TTTGTTTATT GGCCAGGGG TATGTTTGGT AATACTTCAT

```

Figure 8 (Page 60 of 73)

76/162

194401	CAGGCATGAG	TAGTACGTCT	TGGAAGGTGT	GGTCTAAAGC	CTAGACTCCT	ATCTGCTTCC
194461	TTCAGCATT	TCCAGTGTAT	CTGTCTCTG	TCTACCTTAG	GATAGGGGTC	TCCAGAACTT
194521	CCATTACAT	TTAGAAGAGG	GCAGCGGCTT	TCTATGGAAA	ATATGAACTC	TCATTTCATCT
194581	CTATTCCTTC	TTCTAGCTAT	GGTCCAGCTC	AGCTGTTTGG	AATAAAGTAT	CTATATGAAG
194641	TCTGCGAATG	GTTCTCAGAC	TGGTTGAACA	TTAGAATCAC	CTGAGTACCT	TCTAAAATTC
194701	TTATTACCCA	GGGCATATCT	CAGAATGAGT	ACCGCAGGGT	AGGGATAGGA	TTAGGGATCA
194761	TGATCTCTGG	AGTCTGGTTT	AGGCACTAGT	GCTGTTTAAA	ACTACGTTCA	TGAGGTGGAG
194821	GTTGCAGTGA	GCCGAGATGG	CGCCACTGCA	CTCCAACCTG	GGCGACAGAG	TGAGAGTCTG
194881	TCTCAACAAA	ACAAAACAAA	AAAAACCAAC	TACCCTTG TG	ATTTGAATGT	CCATCCAAAA
194941	TTGAGAACCA	TTAGGTAAGG	CCAAGCTGTA	TAATTAAAGA	GCAGTTTTC	TTTGTCTGGT
195001	GTGGTGGCAG	CTTTTGTGATA	AGGGAAGTAT	TGTTGCCATC	CACATACCTG	AGCCTCACTC
195061	CTGAGAACAC	TGGTGTGTAT	GTTGCTAAAA	TTCCCAGGT	GATTCTGAGG	TTCTTCTCTG
195121	GATAAAACC	ACTGACCCTG	GGAATGTACC	CACTGCCAAT	CTCCTGCGTA	AACCTTGGAT
195181	ACTGGGAAGC	CTACAGTTGA	AAATATTGGG	CTTGAGATCC	TGAAACAAAT	CTTGTATTTT
195241	ATTAAGACTA	ATATTGTGTA	CAGTGCAGCA	AATCAAGGGA	ATTTTGGTGG	CTGAGTTCTT
195301	TTAGAACCTT	TGCATTGAAA	TAGGTTCAAG	CAGCAATAAG	TTAAACTAC	AACCTCAGCT
195361	AAAGGATTAA	AAGACACGTG	AGCTGGGTAG	GATGAGGTCT	AAGGTTGGGT	GTGGCGGCTC
195421	ATACCTGTAA	TCCCAGCACT	TGGGAGACT	GAGGTGGGTG	GATCATTGTA	GGTCAGGAGT
195481	TCAAAACCA	CCTGGCCAAC	ATGGTGAAAA	CCCATCTCTA	CTAAGAATAC	AAAAAAATTA
195541	GCTGGGCGAG	GTGCCAGGCA	CCTGTAATCC	CAGCTACTGG	GGAGGCTGAG	GGAGGACAAT
195601	CACTTGAAC	CAGGAGGCAG	AGGTTGTAGT	GAGCTGAGAT	CGCACCCTG	CACTCCAGCC
195661	TGGGTGACAG	AGCAAGACTC	CATTTAAAA	AAAAATAATA	ATAATAACAA	TAATAATAAT
195721	TCAGACATAT	CCAGGCATCA	AACAGATACC	TGGGGCAGAT	GAATAGTCTT	GAGATTCAAG
195781	TCACACATGA	AATTTAGGTG	GAAAATGACA	TTGGAGAAAT	TTGAGATTAT	GATGAATGGA
195841	AATTTTTCAA	AGAGGAATTT	CAGGCTCTGT	TCTTGAGGGG	ATAGATGGAC	TTCCAACAGC
195901	AATAACACAG	GATTAATGAG	GACTTGGGAT	GTTACATAAA	TTAGAGATGT	TAGATGGATA
195961	AAGAGATAAA	AGTACTCTCT	CTAAGAACAT	GGGACCAGAG	ATAGGCTCAC	TTCTAACCAT
196021	CAGATATAAC	TAGCAGACTA	AACGGTCTAA	AAATAAAAT	CATGCCCCAC	TCCTGCTTAA
196081	GACATTTTAA	TTACTCTCAG	TAACTCTTCA	GTTTTTCTAC	TGTGTTATCT	TTAACTACAG
196141	GGTTGGTCTG	GGTGTGCAAC	ACAAGAAAGC	CTGGCATATA	CATGGATTCA	AGTGTATGCC
196201	ATGTGCAGGT	ATTCTTTCAT	GTACTATTTC	ATGTATTCTT	TTTCACATCT	GTTTTTTCCT
196261	TCATTGAAGT	CAATGGCTGA	TATTAGATTC	TACTATTTCAT	GTGTACTAGT	TATATATAAT
196321	TGTTACAAA	CAAATTAGCA	AAAACCTTAGT	GGCTTAAAGC	AACACACATT	TATTATTACC
196381	TAAGGTCTGT	GGATAGAAGT	TCTGACATGG	CTTAACTGGG	TTCCCTGCTT	CAAGCCTCAT
196441	GTGGCTGCAA	TCCAGGTGTT	GGCTGAGTCT	GAATTCTCAT	CAGAGGCTTG	ATTGTGAAA
196501	TTTCCACTTC	CAAGCTCCCT	CAGGTTTGTT	GAAAAATTCA	GTTCTTTGCA	CCGGTAGAAG
196561	CTTCTTGGTA	GAGGCTGATT	CAACTTCTAG	AGGCTGTCTG	CAGTTCCTGT	CACCCAGGGT
196621	GGAGTGCAGT	GGAGCAATCA	TAGCTCACTG	CAGCCTTGAC	CTCCAGAAT	CAATCTGTTC
196681	TCCCACCTCA	GCATCCTGAG	TAGCTGGGAC	CACAAGTGTG	TGCCATCACA	CCTGCCTAAA
196741	AAACAAACAA	ACGAAAAAAA	ACCCCAGAG	AACTTTGTAG	AGACAAGCTG	GTCTGGAAC
196801	CCTGCGCTCA	AGCAATTCTC	CTGCCTTAGC	CTAAAAGTTC	TGGGATTATA	GGTATAAGCC
196861	ACCATACCTG	GCATATGGCA	AGTCTTGAGC	AGGACAAATA	CAGATGATTT	ATGTCTGTCT
196921	TCCATGGTAT	TCTAGGTTAT	TGTTGAGATG	GTCCTCTATT	GTCTTGTTCC	ATCTATTGAT
196981	TAGATAAAAC	GTTGTTCCCT	CTGTTATTTT	TCAACAGTAG	CTTTTATGTG	TCTCTCTTTA
197041	TCTTAAATTT	CTAACCAAAG	AGCTGCTCTT	TTCTTGGTGT	ACTTTACCTT	TGTTGATCC
197101	TTCTTAACCT	CTTCTTGCCC	TCTGGGGCCT	AAGATGAGGG	CTGTTATCAG	ATGTGAGTCT
197161	ATGGGAAAGC	AAGCAAGAGG	TTCTTCAGCC	TCCGTTGAGC	CTTAAATGTC	TAGGTAGAAA
197221	TCAGTCATGG	CCCTTCCAAT	GTGGTACAGA	CCAGATCACA	GAGACAGGGG	TCTCAGCCAA
197281	GGTCTTGTGG	CCTAAGCCTT	ATAGAAATAA	TGAGTGTTTA	CTTACTTGGA	GAACCTCCCTT
197341	GGAATATCTT	TTTTTGTGAA	CCTGAGGCAA	CTTTTGGTGA	TTTCTTGATG	TCTTGGGAAT
197401	CTTGGTCTAG	AGCCATTTCA	ACCCGATTTT	TTTTTCATGTC	AGTGGCATT	TGTGACCAGA
197461	TAGTAAATAA	GTTCTATGAT	GTTCACTCAG	AGAAATACAA	TGACTTATGA	TGCGAAGCTT
197521	CTGTGGTTCA	GCCCTTACTT	CATCTTCATT	CCCTCTTATC	TGCATCTGTC	TCTGTCTTGG
197581	GAACAAAAGT	CTGGCTTCAT	TCTATGACCC	CCACGTTGAG	TTTCTTAGTA	GCACTTACTT

Figure 8 (Page 61 of 73)

SUBSTITUTE SHEET (RULE 26)

77/162

```

197641 TTCAATTAGG AGTGTCTCA CTTCTATCCG TCAGACATAA CTAGCCGACT AAACAGTCTA
197701 AATATAAAAA TCATGTCCTA CTCCTGCTGA AAACATTTTA ATTACTCCCC ATCATTTAAT
197761 TTTTTCTACT GGGTTATCTT TAACTTCAGA GTTGGTCTTG TGTGCAACAC AAGAAAACCT
197821 GGCATATACA TGGATTCAAG TGTATGCCAC GTGCATGTAT TCCTTCATGT ACTATTTTCAT
197881 GTATTCTTTT TCACATCTGT TTTTTCCTCT AAAATTTATT TCCTTTTAAA AATGAAAATT
197941 TTGCATTGGA CTAAATTTGT CAAATTTAGT CAAATTTGTT TAAAACCATT TTTAAAATGT
198001 TTCCCGAAGT TTTGAGTGAA GTTAGTACTT CAGAAAAACT GTTTTGTATT TTTCTGTGA
198061 CCTCAGTGCA CTGCTGTGCA TTTCCATTTT TGGCTCCACA CACATTTGTT TTGAGGAAAT
198121 ATAGGAACGA CAAGATAAAG TTCAAGCTCC TGGACATTGC ATAAAAGACC GTCATGACCT
198181 GGTCTGTGTT ACTTCCCTAG ATTTCCCGCT ATTTCCCTAAG TTGAGATTTT TGGTTGGAT
198241 GCTTTGTGTT TTCCTAAAAT CAAAATAGGT TTTTGCCTTT TATGATTATA CAGTAAATAA
198301 ATGCTATTTG TGTGAAACTT TAAACAATAC AAAAAAACC TAAGGAAGAA AGTCAGATTC
198361 ATCTAAAAAT CCTTGTGGCC AGAATTAAC ACCTTAGTTA CTATTTTCTC TATCTCTCTC
198421 TCTCAATGTA TATTTGGTGT AGGTATAGGG GTGTGTGTAG TGTGTGTGTA TGTATATATC
198481 TGTTTCTATT CCTGTATGTG GATGTGCACA ACGCATCCTG CTTTGTACAC TACAGTACTA
198541 GCATTTTCTT AATGTAATTC AATATTGTTG AAAACATTTT AAAAAAGCTT GTATATATAC
198601 ACACACATAC ACATACATGC ATGTATGTAC ATATACACAT ACAGACAAAA ATGTATCCTA
198661 TGTATATCA CACATGTATA CACACTCACA CATACATAGA GTTTTACATC CATAGTTTAT
198721 AAATGTTGCT TTTTTTTGGT CACCTTTTGG CTAAGTCTTA CACTTTTTTT TTTTTTTTTT
198781 GAGACGGAGT TTTGTTGTCA TTGCCAGGC TTAGTGCAGT AGCGCGATCT CACCTCACTG
198841 CAACCTCGAC CTCCCGGGTT CAAGCGGTTT TCCTGCCTTA GCCTCCTGAG TAGCTGGTAC
198901 TACAGGTGTG CGCCACCATG CCTGGCTAAT TTTTGTAGTT TTTTATAGA GACGAGGTTT
198961 CACCATGTTG GCCAAGCTGG TCTGGAATC CTGACCTCAA GTGATCTGCC TGCCCTCAGAT
199021 TCCCAAAGTT CTGGGATTAC AGATGTGAGC CACTGCACCC GGCCAAGTCT TACACATCTT
199081 TTTTTTACCA CTAACTGTT TACCCAAACC TGATAACCCA AGTCAACAGC TATTATGGCT
199141 CACACAATCT TATGTAAACA AAGATACAGA TATATAGAAT TTTCTTGATT AATATTCAGA
199201 AAAAAATGGA GTCCCTTTAT ACGTCCTTAG TATCTGCTTT ACTCATTTAA AAATGTATTA
199261 CATTATATGA AAGTATTCAG GTCAAATGTT ATAGATGTGA TTCATTCTTT TTAAGTGTGT
199321 TATTTTCTG CAATGACTAT GTATCACAAA GTACTCAGTC TTCCACTGAT GAAAATTTGG
199381 GCTATTTCCA GTTTGTCTTC CATTTTCTTT TCTTCTCTTT GGATTTTTCAC TCAATGTGTT
199441 TACTAATTTA GGAAGAATCA ATAGTTTTTA TGGTATTACT TCTCCCATTC AAGAATATAG
199501 CATATGGTAT AGTATAGTAG AGTACTTAGT TTAATTTAGC CAGATCCTGT TTTCTGCCCT
199561 TTAATAAAAT TCTATCATT TCTGCCTTTG AGTCACATT TCCTTGTTCA TATAATTCTT
199621 AAAAAATGTA TAGTTTTCAT TCTAAGGGAA CATAAAACT TCTTTCCATT TCTATTCCTG
199681 TCTAGTTAAT TCTACTATTG GGAAAAGTAA CTGTTAAAAA AAATCTTAT CTTTCCAGTC
199741 AGTTCACCAC ATTTCCCTTA TACCTTTGTA CTTTAATCCC CAGTCATGTT GAACACTTCT
199801 TATTCCTCAC ACCAAGCCTC AACGGGTTTG CTCTTTCTGG AAGGTGCTTC CCCTGTATTA
199861 CTGACTTATT CATACCACAC ATGGAGACTG GCGCAGCCCT GTTCTGCCTG GGAAGCCTTC
199921 CCCTGATACC CCCAGTTGGC AGGAGTCTTC ATTTGTTCTT TTCTAGTCAC CTGTGCAAGT
199981 TTGTATTGTT CATGTTTATC ATCCTTCATT CTAGTTGTCT GTCTCTGTGT GTGGTCTCAT
200041 TCAGTGGACT CTGAACCTCT ATGAAGTCAT GTCATGGGTC AGATCTTAAT AAATTAATAT
200101 TGTCCGAAGC TAATGTCATG TCTAGAATAC AGAAAAATTA TCAAAAAAAA ATATAGTATG
200161 TTGGCTGGGC GCAGTGGATC AAGCCCCTAA TCCCAGCACT TTGGGAGGCC GAGGCAGGAG
200221 GATCACATGA GGTGAGAAAT TCAAGACCAG CCTGGCCAAA ATGGTGAAAC CTCATCTCTA
200281 CTAAAAATAC AAAAAAGTAG CAGGCGTGTT GGTGCCACAC TGTAATCCCA GCTACTCAGG
200341 AGGCTGAAGC GGGAGGATCA CTTGAACCTG GGAGGCAGAG ATTGCAATGA GCTGAGATCA
200401 TGCCACTGCA CTCCAGCCTG GCGACAGTG AGACTCCATC TCAAAATAAT AATAATAATA
200461 ATAATAATAA TAATAATAAT AATTGTATGG AATTGAACTG CTCTGATTGG AAATAGCTGT
200521 TTTTTAAAAA ATTATTATTT TTTAAGTTCC TGGGTACAAG TACAGGATGT GCAGGTTTGT
200581 TACATAGGTA AACGTGTGCC ATGGTGATTT GCTGCACCTA TCAACCCATC ACCTAGGTAT
200641 TAAGTACAGC ATGCATTAGC TCTTTTACCT AATGTTCTCC CACACCCCA CCCCATCCTC
200701 CCCCACAGG CCCAGTGAG TGTGTTTCCC CTCCCTGTGT CCACATGTTT TCATTGTTCA
200761 GCTCCCACTC ATAAGTGAGA ACATGAGGTG TTTGGTTTTT TGTTCCTGCC TTAGCTGTTA
200821 ATGTCAGGCC AGAGAGGCTT AAATTTTAA GGATCTCTGG ACTTTTCTTC TACATTACTC

```

Figure 8 (Page 62 of 73)

SUBSTITUTE SHEET (RULE 26)

78/162

200881	TTGATGTTTA	TAAATGTTAC	AACCTCTTTA	ATTTCAATTTA	ATGTATACCT	TATTGAGTTG
200941	ATTTAACTGA	GTTAACTTTG	TTATATGAAA	ATCATGATTG	GGAGTGAGGG	GGTTAAACCA
201001	GCTACAGAGA	TCTTGATTGT	TGGTGGTGAA	GCAATGCAAG	AATTCATTCA	TTCAGTAAAC
201061	TAATGTTTAT	TAAGCGTGTA	CTGTCTTAGT	CTGTTCCAGAC	TGCTGTAACA	AAATATCATA
201121	AACTGGGTGA	CTTATAAACA	ACAAAAAATT	TATTTCTTAC	AGTTCTGGAG	GTGGGAAGTC
201181	TAAGATTAAG	GCCCTGGCAA	ATTAGTGTTC	TGGTGAGGAC	AGGTAGCCAT	CTTTTTGCTG
201241	AGTCCTAACA	TGGCAGAAGG	GTTGAATAAA	CTTCCTTGGG	TTTCTTTTAT	AAGGACACTA
201301	ATCCTAGTGA	TGAGGTTTCT	GCCCTCATGG	TATAACTACT	GCCCAAAGAC	CCCTCCTTCT
201361	AATATTATCA	CTTGTGGGT	TAGGATTTCA	ACATGAGTTT	TGAGAGGATA	CAGACATTTG
201421	GATCATAGCA	CACACCATAG	GACAGACACT	GTGCCAAGAA	TTGTGGATAT	AGTGATTCTC
201481	AAAATGAACA	AGATCCCCCTC	AGAGAGCTTG	CAAAATCCAG	CTATAAAATT	ATGCTTTTTA
201541	AACAAATTAT	GCAGTTTGAA	AAATCTACTC	TGAATCTTAC	TTGTGGCATT	GAATACTTTC
201601	GGCCACTCTT	TCCTTATTAT	ATTAAATATT	TACTCTTGTT	TGGGGGATCC	AGTCTCACCT
201661	ACTTTTTCTA	CCAGAACTGG	TATCAGCTCA	TGCTCTGCCT	TATGCAAATT	AAGAAAATAT
201721	CATACCTTTT	GGGTAAATTA	AGCCAAGAAA	GTTCTCCTTT	CTTCTCTTTC	TCTCTTCTT
201781	TCTTTCTCTC	TTTCTCTTTC	TTTCTTCTC	TCTCTCTCTT	TCTTCTTTC	TTTCTTCTT
201841	TCTTTCTTTC	TTTCTTCTT	TTTCTTCTG	ACAGGGTCTT	GCTCTATTGC	CTAGGCTGGA
201901	GTGCAGTGGT	GCAATCTCAG	CTCACTGCAG	CCTTGAAGTC	CAGGGCTCAA	GCAATCCTCC
201961	TGAGTAGCTG	GGACTATAGG	CATGTGCCAC	AACATCAAGC	TAATTTTTGC	ATTTTTTTGT
202021	GGAGACGGGA	TCTCCCTATG	TTGCTAAGGC	TGGTCTTGGA	TTCTTGGGCT	TATGCGATTG
202081	TCCTGCCTCA	GCCTCCCAA	GTCTGGGAT	TACAGGCATG	AGCCACTGCC	CCTGGCCATT
202141	ATAACTATTT	TCATTGGCTT	ATCAGGCACA	TGATAACTAT	AATAAATCAA	TAACCAGAAT
202201	TTTAAATATA	AGAAAGGAAG	GAATGTGTTT	AACCTCTTCT	GCTACCCCTC	TATCCCTCAA
202261	AAGGGTAGGC	TGAATGTTGT	CCTCCAAAGA	TATCCATGTC	CTAATCCCCA	GAACCTGTAA
202321	ATATATTACC	TTATATGACA	AAAGGGACTT	TACATGTTTA	ATAAGTTAAG	AATTTTGAGA
202381	TGGGCAGATT	TTCCTGAATT	TTGCAGATGG	GCCCTAGTGT	AATCACAAGG	GTCTTTATAA
202441	GAGACAGGCA	GAAGAGTCAG	AATAAGAGAA	AAATACTTCA	AGATGTTACA	CTGCTGGCTT
202501	TAAGGTGGAG	GAAAGGCCAA	GAGCCAAAAA	ATGCAGTGGT	CACATCAAGC	TGAAAAAGAAA
202561	AAGAAATGGA	TTTTCCCTTA	AAGCCTCTGG	AGGGGGCACA	ACCTTGCCAA	TACCTTGATT
202621	TTGGCTCAGT	GAAACCCATT	TTGGACTTCT	GACCTTTAGA	ATTGTAAATA	AATAAATAAT
202681	TTTGTTGTTG	TTCAAGCCAT	TACAGTTGTG	GTAATTTACT	ACAACAGCAA	TAAAATAGAA
202741	TTAAATACAG	AGATCTGAGG	AGTTGAGTAG	GATAAGCCTA	CTCCAGCAGG	TTATTTTCGGG
202801	AGTATGGTGA	GACTCACTAG	GATGGCGGAA	CTCAATTAAG	GAAGTCTGAA	GCTGATAAGC
202861	CAGAGAGGGA	AGGCTCTCAT	TTCAATTTAT	AAGGGTTGCG	TCACACTAGG	AAGATCCAAT
202921	AGCAACCACA	GTCTCAAAAT	TAATGATTAC	AAATAGGACA	CAATTCCAAG	AGTCGGGAGC
202981	CAAGCAGAAA	ATGGATTAGG	GAAGACATGG	ATGATATGAA	ACAGGAAGGA	GGGGTACAAG
203041	GCAGCTTCCT	GGGAAGTTGC	CAGGGCAGTC	ACAGTTCACA	TTCATTAGGC	TGTGGGCACC
203101	AAATGCATAT	GGAAAATCTA	GCTGACTTAA	CTGAATCCTT	GAAGAGGAAT	GAACACCTCA
203161	TTTATTGAGG	AGCTACTACC	AATTAGAATA	TGTATTTTCT	TTGTTCAATA	ACCCCATGAG
203221	TACAGTAACA	CAATCCTTGC	TTTACTAAAG	CGGAAGCCAA	TTCAAAGAGG	TTCACTGACT
203281	TGTCCAAGCT	CAGGGAAAAC	ACTAGGAAGT	GAATATGGGT	CTGACTCCAT	CACTGATTTC
203341	AGGAGCCCTG	CCCTTTCCTC	CACACCATGC	CCCCTTGCTT	TCAGAAAAAA	AGGCTTGTTG
203401	ACTGAATGGT	TGTATGCACA	GTTCAAAGCA	GAAACACACG	ATGACATCTT	TTGAGATACT
203461	CTAACAGTGA	GAACTTGAAA	ATGAAGTTAA	AAATTAAGCG	GCAAAACCAA	GCCGAGGCTT
203521	TCTGAGAAAAG	TGGGGCCAAA	CCTGTTGCCG	TCTGACTGCC	ACGTGGCTCA	CTATTTATCC
203581	CTGTAAAAAT	CTGCAAAAGT	ATTGAAAAGG	GAAGAAGGGA	CAGAAAACTC	CCTCCTTTTC
203641	CAAGTTAGCC	TTATAGTCTA	GGGCTTAAAA	TACTGGTTTA	ATGGTGAAAG	TAAGTGCTTT
203701	TCTTCTTTTT	GGGTAGAAGG	ATTATTACTA	ACTTACCAAA	GGTCCATTAA	GGGGAGGGAA
203761	CAGTTTTTAGG	AGAAAGTCAGA	GAAAAGACAT	TAACAGCAAC	ATAAGGATCT	CCATCTGGTA
203821	ATATTGCCTA	ATTCCAAAAT	GAAGAGACTC	TCTGAAAAAG	ATAACTGATT	CAATGAAGAC
203881	CCTAGGGCAA	GGCTTGAGAA	GCCACTGGTA	CCAATGGACA	CTGTGGACAA	TGGTCATTTC
203941	TCCAAGGACG	CTGTGAGTAT	TAAGTGTGAT	GCTGTGATTA	GTGAGACTGG	GATTGGCTGT
204001	GGAATGAAAT	ACTGATCAGA	ACTGACAAGA	TTGTGTTTGG	GGACTGTGGC	TAACGAGTCT
204061	TTTCAGACTT	CTATATGAAT	TTGAAATGGT	CTCTCAGGAA	AAGGAGAACA	TGGCCGGGCC

Figure 8 (Page 63 of 73)

79/162

204121	TGGTGGCTCA	CGCCTGTAAT	CCCAGCACTT	TGGCAGGCTG	AGGCGGGCAG	ATCACTTGAG
204181	GTCAGGAGTT	TGAGACCAGC	CTGGCCAACA	TGGTGAAACC	CTGTCTCCAC	TAAAAATACA
204241	AAAATTAGCA	GGGCGTAGCG	GCGCGTGCAC	CTATGCGCAT	GCATAGTGCG	CGTGCCAGCT
204301	ATTGAGAAGG	CTGAGGCGAG	AGAATTGCTT	GAACCCAGGA	CGTAGAGGTT	GCAGTAGTTG
204361	AGATCATACC	ACTGCACTCC	AGCCTAGGTG	ACAGAGTAAG	ACTCTGTCTC	AAAAAATAA
204421	TAATAATAAA	AGAAAAGGAG	AACATGACCA	AAGTTATGAA	TAAGACTGAA	GGCAAGAAAA
204481	TTGTACGCTT	GTAGAGATCA	CCTAGCTTGT	TGCCCTCATT	GTACAGCTAA	GAAAAGGCAC
204541	CCAGGGACAT	TGTGGTCAGC	ACCAATTTCT	CAGAAAGATA	GGCAGATGAT	GAGAGGGCCC
204601	TCAGTTTTC	TAACACTGAA	GGAATTGCTT	CTATGTTTTC	TGGTGAACCT	CTCCCCACTC
204661	ATCTTGAGGA	TTCCAGGCCA	GAAGAATCCA	CTTTAAAAAA	GAAACATTTA	AAACCAATTT
204721	AACAACCAAT	CAAAGGCACT	TTTATAGAAA	TACATTTTCAT	TGCTGTAGG	CCTGTATTTA
204781	TGGATCTGAG	AGGGCTAGAC	TGCCAATATT	GTGACTGTTT	ATTATTATTG	CTGTTGCTAG
204841	TATCTAGAAT	ATTATACAAC	ATATAACACT	TTGCAATTTA	CGAGGCATGT	CTCATACTTT
204901	TGTTTTCACT	CCAAACTGCC	CAGTGAAGTA	ACATTATCCC	AATTCTTCCT	ATGAAACAGT
204961	GAAAGCCCTA	AGAGTTTTTG	AAACTTTACC	TGGTTTTACTC	AATTTGGGAA	TGGCAGAGCA
205021	GAATTGAGTC	CTTGAATATC	CTCCCACTGC	AGGTTTCATGC	TCTTTGATCT	AGGTGTAACA
205081	TTTACTCTGA	GTAACCTAGG	ACTCTGGGCT	AACAGAGATG	AAGCAAGACA	GGCTGGATAT
205141	TAGGAGAAAT	TAAGAGCAAT	CTAACGACCA	TTATAATAAA	ATCATGAGTT	CTAGACTTAA
205201	AAAAAGGGAA	AAACCTGTTT	TTTGCTTTAT	GCGTATACCA	TAATATTTAC	ATTATTTATT
205261	TTTTTCTCAA	ATTCAACCTA	TACTGTGTCA	AGTAATTTTT	TTAATATAAA	CATTTTCCTT
205321	TAACTTAATT	TCAATTCATT	TTTCTGTGTC	TACTTACAAC	TTTGGCACTA	GAATTCACAA
205381	TTTTTTTTTA	GAGGTATATC	TCCTTAAAGG	GAAGGGTTCT	GACACTGTTA	CATGTTCTCA
205441	ATTGTTTGCA	AATAGGTTAA	TAATTATTCC	AGTGTCTCTA	AGTACATATC	AACCATGCCA
205501	GTGTTGAGCC	TCCATAATTT	TATTAGCTTC	TGTGCTTATT	TTGGAAAAAC	ATTTCCCAT
205561	ACCATGAAAG	ACCTCAGTTT	AGGATGGTTT	GGTATGTTAG	CCTGATTTCT	GCATTGCTCT
205621	CATGCAAAGG	AAAATAGGAA	ACGAAGAATC	GAAATTACCT	ATTGATACAA	AATCAAAGTA
205681	GCATTTGAAA	CCATAAAACT	TAAGTAGGGC	TTTTCATCCT	TTCTCGTTAG	ACAGCAACAG
205741	AGAATGGGAA	GAAAAACTAA	AGTGATGGGT	TTGTGATACA	ATTCCAGTAA	CATAAGAGC
205801	AAGGAGAAGT	AGTTTTGTTG	TGTTTATGTT	TAATATTCAA	AGCTCAACCT	AAAAGTATTT
205861	TTCAATTATCA	AACTTCCTTC	TAGAATAAAT	GATTAAAACT	TGATTTAAAA	TATACAAATT
205921	CTCCTTTATA	ATACCTCAAA	ATGGAGCTAC	CCCATTGAGT	TTTAAGCTTG	TGATTTAAAT
205981	ATTACGAAAA	CAAAGGGGAA	GTTGTAATAG	GTAGAACAAG	CAGTAGTCTA	GGCATTAGGG
206041	GATCTGGTGC	TGGCTCTGTG	CATCATGTGG	TTTCAGGCAG	CTTTTCAAAT	TTTCTACGCA
206101	AATTTTCTTA	TCAATAAAAT	AAACAGTTGG	GCCAGAGGAT	CTCTGAGTCT	CTTTCAGCTT
206161	TCAGTGTTTA	TAAGATTGGA	GAGTTGGTGG	GGAAAGCTTT	AAGTGGAGTG	TAAGTAATTG
206221	CAGCTGCATG	TACAGTTAAA	GAGTTGCCCT	CAGCCAAGCC	ACGGGATCTT	GCATAAAAAA
206281	TGAAATCAAA	TAGAAAATGG	TCCAAACTCT	GGGTTTGACC	ACAGATGACT	CTCAGTAGGA
206341	TCTGAGTGTA	GAGCAATGAG	CTGAACTCCT	GATATCCAGA	TGTTAGCAAG	ACTTGGAGGC
206401	CTTCTAAGGC	AGAGCAACAA	CCAGTATCTG	TCCTGGTGCT	GACCTGATCT	TACTAGCAAT
206461	TGGGCCTCCA	TTTGGGTCCA	TTGTACAAAA	CAACAACAAC	AACAACAATA	AAATCTCCAA
206521	ACACCCAAAA	TTCAAAATTT	AGATGGAGAG	ATACTATTCC	CAGAATTCTA	GAGATATTTG
206581	GAAAGCAGAA	AACTATACTT	GCCATGCTGA	TGAAGTCCAA	TTATTGCTCT	TTTAAATACA
206641	TTTAGCTACT	TCTGAATATA	AAATGAGTAT	CTACTAATTA	TTTACAAAAT	CACTTGGTAA
206701	ATATAGAAAG	TCACAAAGAA	TGAAGTGATC	ATCCTGTTTT	GTAACCCAGA	AATAGTCATT
206761	ACTGGCACTT	GTGTGAATCA	GTTTCTATTG	CTGTATGTGG	ATGTGCACAG	CGTATCCTGC
206821	TTTGTACACT	AGAGTACTAG	CATTTTTCTA	ATGTAATTCA	ATATTGTCTG	AAACATTTTA
206881	AAATAGCTTC	CATCACAATA	ATCTATCAAA	TTGACTTGCC	AGACTCTCAT	TATTAGGTTA
206941	ATTTATCTCT	AACATTATGC	AGTCATGAGT	AATACTACAA	AGGATATTTT	TGGACACAAT
207001	TTTTCATCTA	TGCCTTTCTT	TATAATCCTT	CATCCTAAGG	TCACAGATTA	TGAATATCTT
207061	TAAAGTACGG	ACAAGTCTTT	TAAATTTTGT	GTGCAAAAAC	AGTGCAAAGC	CTTGAATGAT
207121	AAAATAGAGG	TTTGATATAT	GTGTTTTTTT	GTTTGTGTTT	TTTGAGACGG	ATTCTGCTC
207181	TGTCCCCCAA	GCTGTAGTGC	AGTGGCACGA	TCTTGGCTCA	CTGCAACCTT	TGCCTCTTGG
207241	GTTCAAGCAA	TTATCCTGCC	TCAGCCTCCT	TAGTAGCAGG	GTCTACAGGC	ATGTGCCACC
207301	ACACCCGGCT	GTTTTTGTAT	TTTTAGTAGA	GATGGGGTTT	CACCATGTTG	GCCAGGATGA

Figure 8 (Page 64 of 73)

SUBSTITUTE SHEET (RULE 26)

80/162

207361 TCTCGAACAC CTGACCTCAA GTGATCCACC CACCTCAGTC TCCCAAAGTG CTGGGATTAC
207421 AGGTGTGAGC CACTGCACCC GGCCGATACA TGTGTTTTTA AAGTCACAGA AATTTTCAGAT
207481 GTCTTGAAGG ATTTTAAGCA ATTTAAAAA TAAAGTCATA GAAGCTTCAA TTTAGGAATG
207541 AATGGAAAAT TGATGATATT CTTAGGATAT GGATTTTTCC TAAAAGAAAC AAATGTATGC
207601 ATCCCCAAAG ATAATTTGAT TAGTATACAA ATATTAAAT AAACATGTCC ATATTTAGAG
207661 CCATGAATTC TCTTTGCCTG TCACAATAGC TGGATTATAT CACAATTGTA GTAATTAGTC
207721 CCTGTTCAAT ATAATTTTCT AGGTGATATG AAGACTTTGT CAGTCCAAGC AAGTGTCCAC
207781 ATTGTGTGTA GCAAACATGA GAATAACAT TTTAACTTT TAAATGTAAT ACATATTAGT
207841 GTTATGTAAT GTCATCCTTC ATGTTCGAAG GCACATGGAA CATTGTTCTG GTGGTACAGA
207901 GGGGAGAGAA ACACCATCAG AATGAAAGGA AAGACCGCTC TGGAACCTTC CTCCTTAGCT
207961 CTTGAGCTTA GTTTAATTGT CCTGTCTTAT GGTCTGCTAC AAGCAATACC ACTCTTCACC
208021 TTCGCATGCT TCTCTGTGGT TTGATAAAGT ACATGCAATT TTTCATTTAA TTCTTCAGC
208081 TGCATAAGA AAGGAGCCTT ATCTTTATTG AACAGATGAG GAAATGAATG ATTAGAGAAT
208141 TTAAATGACT AGCTCTAGGT CACACAGCTG GAACCTACAG CCAGATTTCC TTTTAACAAT
208201 CCTGTAACCA AAAGCATACC AGTAGTGCCC CATAAAATGT AAGTTATAGA GCTGTGTTGG
208261 GTCAAAACCTT TTAGTGATGC TAAGAGGAGG CAACATTAAC AAGGGGAAAT TATTTGTGTA
208321 TTATGTTTTG GATTATGTTT TCTCCATAGA TAAAAGACTG TCGTAGTAAA AGAGATTACG
208381 GGCACAGGGA AACTCCACCA CAAGCGTGG TACCATTTC CACAGAAGCT AAATGGACGG
208441 GAAGCCTGCC ACCAGGAAAG GTAAAGCCAC TGCTCTGTT TGCAGGCTAT GTTAATAAGC
208501 TGAAGCTTAT TCCGACACAT TTACACATCT CTGCATCACA CTGACCTTC GTAAAGATAC
208561 TCCCAGTGTA ACATTGGAGC CAGCTCCAGC CCCTGATCCT GTTGCTTTTT CCTTAGCCCC
208621 ATGAAATCAT CTGTGAGAAA TTAAGCCAAA TAAGCAATAA ATCCTGGGAT CTAGGGAGTG
208681 GAATAAGTTT TGGGAAAGTC TTTTTTTTTT TTTTTTTTGA CTGAGTCTTG CTCTGTCTCA
208741 CAGGCTGGAG TGCAGTGGTG CGATCTCGGC TCACTGCAAC CTCTGCCTCC CGGGTTCAAG
208801 TGATTCTCCT GCCTCAGCCT CCCGAGTAGC TTGGACTACA GGCACACACC ACCATGCCCCA
208861 GATGAATTTT TGTATTTTTA GTAGAGATGG AGTTTCGCCG TGTTAGCCAG GATGGTCTCG
208921 ATCTCTGAC CTCGTGATCC ACCGGCTCG GCCTCCCAA GTGCTGGGAT TACAGGCATG
208981 GGCCACCACG CCTGGCCCCG GAAAGTCATT TTAACCAAC CTATGTATGA ATCCCTACTA
209041 TAATATTCTC ACCAAGCGGC TGGCTCTTTC TCCTGAGCTT GGAAACCTCC AGTAAAATGG
209101 AAATAATTAT TTCCAGACC ACCACTCTTA TCTGTGAGCT TTTTGGCCA TTAATAATTA
209161 TTTCTTCCAT TATATTTTTA TCTGTGCTT CACAGGTTTT CTCTTTCTTT CACTTTAGTG
209221 CTTTTCTTCA AATAAGCAGG AAAAATCCAA TCTATCATGC ACATGGGAAC CCTTTCAATA
209281 TTGGTCTGTG GTTGTTCCAT TTTATGGGGA TGCTTTTAAA GAAAAAATT GTCTTTTCAA
209341 TATATTGAAT ATCTTCCAGC ACCACATCAC CTGCAAGCTT TGTAAAAATA GTTCTACATA
209401 TTAATTTTTT TTTTTTTTTT GAGATTGAGT CTCATTCTGT CACCCAGGCT GGAGTACAGT
209461 GACATGATCT TGGCTCATG CAACCTCTGC CTCCTGGGTT CAAGTGATTC TCCTGACTCA
209521 GCCTCCCGAG TAGCTGGGAT TACAGGCATG CATCACCATG CCTGGGTAAT TTTTGTATT
209581 TTAGTAGAGA TGGGGTTTCA CCATGTTGAC CAGGCTGGTC TCAAACCTCT GACCTCAAGT
209641 GATCCACCTG CTTAGCCTC CCAAATGCT GGGACTACAG GCGTGAGCCA CTGCACCCCA
209701 CGTAGTTTTT TTTTTTTTTT AAGTTGAACA TATGTGAAGG CAGGACCTAG TGACACATAG
209761 CAATAACATT TCCAAGTAGA CATTACACTA GGGAAATAGT CGAAGTGCTC ATTTAAAGTA
209821 CCATCTCTCA AATGTATTAA AAGAGAATCC TTGGATGTGC AATACCTTAA TTCAAAGGCA
209881 GCTCGTTATG TATAAATCT CAAGCTTTGT GATAAACAAA TGTGCATAAC AGATGGGACT
209941 ATTCACCTAC AGCCAGGGA ATTTTATTGA CGCTGAGAAG GTTATGTGAC TGGCTCTGCC
210001 ACTGTCAATC CCATTCATCT CATTTTGGAG CAATATGACA TAAATGCCTT ACATGTGGGT
210061 TTTCTCTATT TATCATGTGT TTCCTATCCC CTTGAAAGAT GGCCATATTT GCTTTACTTG
210121 GTTATAAGAT CCCATATTCG CTGTCTTGAA GCCAACCAAA TAATTGACA AAGTGGGTTT
210181 GTAGTGCTGG CTATTTTGGT GAAAAAAGA CAATGAGACT TCATGTGTCA TCCAAAGTTC
210241 TATCAGATCG AGCTGTGAGA GAAAGGAAA GAAAGGGGTC TCAGTCAGGA TGCTCACTAC
210301 ATACATCTGT GTTGTGTCT AGGTCCAGAT TTCTGTTTAT TACGCTATGG GCTGGCTCTT
210361 ATCATGCACT TCTCAAACCT CACCATGATA ACGCAGCGTG TGAGTCTGAG CATTGCGATC
210421 ATCGCCATGG TGAACACCAC TCAGCAGCAA GGTCTATCTA ATGCCTCCAC TGAGGGGCCT
210481 GTTGCAATG CTTCAATAA CTCCAGCATA TCCATCAAGG AATTTGATAC AAAGGTAAGT
210541 ATGATGGAAG ATAGGGCTCT TTGTTGAGAG AAAAACTTT GAAAGGAAG CATAGATCTT

Figure 8 (Page 65 of 73)

SUBSTITUTE SHEET (RULE 26)

81/162

210601 GATTCTGTGG AGTATGGAAG TATACATTTT CAATGACAAA TAAAACTGA CTGGAECTAT
210661 TTTTCTTTGA GACATTGCTT ACTTCAATAA TAAAAATAAG ATTTTCATTGA GGTATTATATG
210721 ATTATAAGGT GGGGGAAGTG TAGAGTTAAA TGTGAAAAAT TTAATAATGG AACAGTTTAT
210781 GTGATGTCTT CAATGAAAAA CTAGGTATTA CCTGGGCACA TTCTTATAGG TTACTCAATC
210841 CTATTCAGTT CTCTGCCTGT TTTATTGTTT CTGAGCAATT TTATATCCCT GTAAATTCTA
210901 TATAACCAAT AGAAATGCAA ACGATTCTTG TCCATAGCTT TGCAAATAAA TTTTGCCAAG
210961 AGAAAAATCA GTTAAAACTT TTCTCCACTC ACCTCCAGT TGAATTAGCC AATTTTGCTG
211021 TTTGTTTGTG TGTTTGTGTT TTGAGATAGA GTCTTCCTCT GTCATTCAGG CTGGAGTGCA
211081 GTGGCATGAT CTCAGCTCAC TGCAGCTCC GCCTCCCGGG TTCAAGAGAT TTTCTGTCT
211141 CGGCCTCCCA AGTAGCTGGG AGTAAGGGGG CATGCCACCG CGGCTGGCTA ATTTTGTAT
211201 TTTTAGTAGA GACAGGGTTT CACTAGGCTG GTCTCGAAT CCTGACCTCA GGTGATCCAC
211261 CCGCCTCGGC CTCCCAAAGT GTTGGGATTA CAGGTGTGAG CCACTGTGCC AGGCTCTGCT
211321 GTATATTTAA AGTCTATTTT AGCATTGCTT CCTGCTTGTG TTATGCGTGA TTCTTTGAGT
211381 TTTCTTTTGA ACCAGTTATA ACATCTTACT TACTTCCTCC ATTAATCAAT GAGTTAAATA
211441 AAATCTTTGT TGTATGTTTA TTTTACATTT ATATGAAAC CATGAATTTA CCCAATTA
211501 AAAATTATCC TTTAAATTAT CTGTACTGT ACATTTCCCA TGTCAATCCCT ATAATTCATG
211561 ATTAATGATT TTATTACATT GGACCTAGCT TATTTACAAT GAGTACATAA ATTTATTGTC
211621 TCCAGTCTTT CCTCCATTAT CCCGTCTACA TATCCACACT GAGTAGATTC ACTACTCAGG
211681 AATCTTGGAC ACCTTCAAGT TGCCAAACAT GCAGTGTTC A CTGGACATGC TGTGTTCTT
211741 CAGAATTTGG GCCTGCTTCT CAGCACACTC ACATCTGCTA TCAATGACCC ATGGAAGTT
211801 TTTGCCCTGA GCAAGCCAGA GTCCCTGTTA GTTTCTTCCA AATGCTACAA GTTCACTTTT
211861 GCTATTTTTT CCGATGAGAT AAAATTTTCC TTTTGTGACT TCTACAAATC ATAGTCAATT
211921 TTCAAGGGAT AGTTCAAGTA TTGCTTCCTT TCTGGGACCT TCCCAAATTA TTAATTTCTC
211981 CTCTCAAAGT CTCTGTTTTA TTTATGTTCA TCCTCAAATC TTGATTCTCA CATGAATCAT
212041 ATACCTTGTA TTATTTATAG TTTTGTGAG TGGGTAAAAT ATTTTCATAT TTATTTCTT
212101 TGGCTCTCTA CTTTATAGCA TGATGCCAGA TATTTAGGGG CCTTATTGCA TTTATTTTTT
212161 ATTTTATTTT AAAATCTATT TTATTTTTTA TTTATTTATT TTAATCTA TTTATTTTTA
212221 GGTAATATTT CAGGTAATAT AATTTATGTA ATTATTTAGG AATTTTAGGT AGTTATTTTA
212281 AAATAATTCA AATTATTTAT TGAGTTATAT CAGAAGAAATG TGATCTTATT CATTTGTAAT
212341 ATGTGTTTTA GGAAGCTCAGT TCAGCCAGGG CAGACCATGA TTCCCAAATC TGACTTTTCT
212401 TTTTAATTAG GCACTGATTT TGGTTAAGAG TTCAGTAAAG TTTTGTGTGT GTGTTTTTAA
212461 AAATTCCTTG ATATAAGAGT CAAGATGTTA CTCAACTTTT ACTAGAAGCA AAATAGAGGA
212521 AGTGCTTTCA CAGATGAAAT ATCTCTCAAT GTTTTCTTCC ATTTACTTCT TCCTATTATT
212581 CATCTATATA ATCATTTTCT TTACCTCTTT TCTTCATTTT TCTGTTTTT CTCTCCTTCT
212641 ACTAAGACAA GCAATTAGG GGTATAATTG GTTATTTGGG AAGGTAGGAA GAATATAGAG
212701 AGAAACAAAA ATCAATATTT TATACTAGGG TCTCACTAAC CTCAAGCAAC TCTGACTGTA
212761 AAGTAGATTT TCATAATAGG ACTTCTTGAC AAAGAGTTTT CCTATTTTTC CCCCAGGCCT
212821 CTGTGTATCA ATGGAGCCCA GAAAGCTCAGG GTATCATCTT TAGCTCCATC AACTATGGGA
212881 TAATACTGAC TCTGATCCCA AGTGGATATT TAGCAGGGAT ATTTGGAGCA AAAAAAATGC
212941 TTGGTGCTGG TTTGCTGATC TCTTCCCTTC TCACCTCTT TACACCACTG GCTGCTGACT
213001 TCGGAGTGAT TTTGGTCAATC ATGGTTCGGA CAGTCCAGGG CTTGGCCAGG GTATCCAGAT
213061 ACTTTCTCAT TCTTGGTGGG ATCCAGATTT CTGAATTCTA CAAAATATCA AAGGTCTTAA
213121 TGATTTTCAT TTCAGGGAAT GGCATGGACA GGTCAAGTTA CTATTTGGGC AAAGTGGGCT
213181 CCTCCACTTG AACGAAGCAA GCTCACCACC ATTGCAGGAT CAGGTAAGTG TGCACAGATG
213241 GGTCATAGCT TTGTCATCTG TTCCATCCCA CTGTGTCTTA TCTTCTATGA ATCAATGGT
213301 TTGGGGAAGA GAGAGAAAAA GTACTGCTGA AAAATTCAAC AATATAAGAC ACTTGCATCA
213361 CAAATAGGAA AGATGCATCT GTGCAGTAAA GACATTGAAG CTTAGAAGTA GAAAAACCA
213421 TTGTGAGCTA GGTTTCAGCT CAGAAAAGCC TTAGTAGTCA GAAAAGCCTT AGTAGTCAGA
213481 AAAGCCTTGT CGGAAAAAGT TTAAACCTTT AAGAATTGCA CACATGGAAA AAGATCAAGT
213541 AAGCTATATA TACACCATCT TAGCAATGAT TTTGAAGTGA GAATTAAGGC TACCACAGCT
213601 CCAGGTGGTA AGGAGAGAAA TCAGGCTGGA AGAGTTGAA GTTTCTGTAT TATTCTAAGC
213661 TCTTTACTAT TCTATTATGA GCTCATTAAT TCTCACAACA ACCCTCTCAT ATAAGTACCA
213721 TTTTAAATTC TTATTTTACA GAGAAGGGAG TTAAGGAAGG TGGAGATTAA GAAAATTGCC
213781 CAAATACAAA TAGCCAGCAG GTGGTAGGTC TGAGATTAA GCCCATGCAG ATTTTAGCCC

Figure 8 (Page 66 of 73)

82/162

213841	CAGAGCAGAC	ATTCTCAATC	ACTATGCTAG	ACTGCCTTTC	CATGGTATGT	GATCCTACTC
213901	AGGCCTCTAC	AGCTTTATCA	TTGCTGTTCT	CCCCAGCCTG	TCGTGCTGAG	AGTATATACT
213961	CGAAGAGCAG	AACTAAAATT	CCATCCAGCT	TCTCACTCCT	AGGTCCACTA	CACAGCTGCA
214021	TCCTGCAGAC	TTTTACCTCA	AGCAACCCTC	CTGCGTTCTT	GCTTCCTTCC	ATCATAGTTG
214081	TAACCATCTC	CTCTATTTGC	AAATACTATC	TGCTGATCTC	TCTCTTCTAG	ACTGGTTTTCT
214141	TTCAACCTTC	TTCCCACCAA	AACCAAGTTA	GCTTGCTAAA	ATAAAGATGG	CACATTTTTTA
214201	CTCACCCGCT	TGAGAATTTT	CAATGTGTTT	CTTCATGCTT	ACAGAGTAAA	GCCTGACCTC
214261	TTTATTGCAT	GAATACAAAA	GTTCTTAGCC	ATCTGGCCCC	AACCTTGTTT	CACTCAACTC
214321	CCCTGTGCAA	GCATGGCTCC	AGTGGCACTG	GACATTGGCT	GCTCTCCACA	TAGATCTGCA
214381	CTGCACTTCC	CTCTGGCTCT	GCTCCCCTTA	GTTTATATGC	CTGGAAAGTT	CTTTGCCCTT
214441	GTTCTTGTG	CCAAAATTCC	ATCTATCCTA	TTGCATAGCT	TATGTAAAAA	CTTCCTAAAC
214501	CTTTTTTTTT	TTTTTTTTTT	TTTTTTTTTG	AGACGGTGTC	TCACTCTTTC	GCCCAGGCCG
214561	GACTGCAGTA	GCGCTATCTC	GGCTCACTGC	AAGCTCCGCC	TCCCAGGTTT	ACGCCATTTT
214621	CCTGCCTCAG	CCTCCCGAGT	AGCTGGGACT	ACAGGCGCCT	GCCACCATGA	CCGGCTAATT
214681	TTTTGTATTT	TTAGTAGAGA	CGGGGTTTCA	AGCCAGGATG	GTCTCAATCT	CCTGACCTCG
214741	TGATCCGCC	GCCTCGGCCT	CCCAAAGTTC	TGGGATTACA	GGCGTGAGCC	ACCGCGCCCC
214801	GCCAAAACCT	CCTAAATCTT	ATAATTATTA	TCAATTTATC	CTCAGATATA	CTTCCACGTA
214861	CATTGTAGTT	TTATTATATT	TATATTTTAC	ATCTTTTTTT	TCAAATTTCA	GTTTGGGACC
214921	CATTAGTGAG	TCATAAAATC	CATTGAGCGG	GTTAAAATCA	TTATTTTAAA	AAATGAATAG
214981	AATAGAATAG	AAATTGTTGG	AGTGCATTGG	ACATGGTAAA	GTTAAATATC	GATTTCATGAA
215041	ACCATCGTTT	GAGGCATATG	TGTGTGGTTG	TATGTACAAG	TGTTTATGCA	TATTGGTGTG
215101	TGTGTTATGT	TACCCTGTAA	AATGCATTTT	TTACTATAGG	TCTCTGTGAA	ATATGTGTCT
215161	TGTTGTTTTT	TAATGTAGAC	TTCCAAAGCC	TACATGGCAT	TTCACTAGTG	ACAATCAATT
215221	TTATTCACAT	TTTTCTCTCC	AATTGGACCA	GAAGCTCTTT	GAGGGCAGGG	GCTGTATCTT
215281	ACCGATTTTT	GTAAGTCTTT	ATTTCTCTGC	CCCTAGCCTC	ATATTAGATC	ATGCAAGAAT
215341	GCAACTGTAA	TCACAAGAAA	CATTCTAATG	GCTGTGATAG	CAGAGAGTTA	CTGTGACAAA
215401	CTAAGGGATT	TAGATTTGGT	CACATTGGTG	TTGAGGAGCC	ATTGAAGAAT	CAGAGAGTGT
215461	GTTACTATTA	TTTGTTAATT	TTAATTATAT	CATATTACTT	TACTGGGGAA	AATCTGTGAG
215521	CTATTTTAGA	AATAAATACT	CTCATTGCCC	AATAATTCTA	AGTCTGCCAC	CTCACTGTTG
215581	GGACATTGTT	TAGGGAGGCC	ACGAAGTCTC	AGCCTTTGAT	ATTTTCATAA	GTGTTTTTCT
215641	CCCTTTTTCC	TTTAGGGTCA	GCATTGCGAT	CCTTCATCAT	CCTCTGTGTG	GGGGGACTAA
215701	TCTCACAGGC	CTTGAGCTGG	CCTTTTATCT	TCTACATCTT	TGGTGAGTCA	CTTTCTCTTA
215761	AATCCTAACG	CCTCCATTTT	CTGAGCATCC	ATTTTGCCAC	CTACACCACC	CACATTCTTC
215821	CTATATGAAA	GAAAAATGTC	TTTATCAAT	GGAAGATGAT	AAAAAATGTC	AACGGTTGGT
215881	ATCATTTTTA	ATCTAGTCAC	ACAACCTTTC	TAACACCTTC	CTGGTGGTTT	TGGGAAGCCA
215941	CACGCACAAG	GTAGAGGAGT	TGACTATTCA	CATGGCACCC	ACCGACTTGT	GATGCAGTCT
216001	TGTCCTTCCA	TATCAAGCAC	CTTCTGCAGA	ATCTCTACCA	CCACATCTGA	AGTGCCTGCT
216061	ATATGCAGTT	AAGATGTCAA	AGATAGTGAA	GTACATTTTC	AATGTGTCTT	CATATTTTCAT
216121	TATAATTATT	ATTTCTGTCC	AAGATGCCTT	TCACCTGTTT	TCTACCAAGT	TAATCTTGCA
216181	AAGTTCAATT	CAAATGTTCC	CTTCCCCATG	GGCCCTTCCA	GGGCTTACCC	TATCAGATTC
216241	TGGCATTCTC	TCCTTTATGA	TATTTCTCTT	CTAGGTTATG	TTGGTGTGTA	ATTATTTTAT
216301	TCTCCTTTTC	TTTCCACTAG	ACTGTGAAAT	GCTTGAGGCA	AGGAATCCAT	TCTATGTTTT
216361	CATCACTTGG	GTGTCATCAT	GGTGCCGTAT	TTTGTAGCTT	AAAATAAAAG	AATCAGTGAA
216421	TCCAGTAAAT	AGAGGGGATT	TAAAGAAAAA	TAGTCCTCAG	AATCTTTTAA	CATAGAATGT
216481	TCTTCAAATA	AGGAATTCCA	ATAATAAGAC	AATTTTCTAC	ACTTGATTTT	GTTTTTATAG
216541	CCAAATGGTG	TCATTAAATA	TAGTCTGGC	CTGAATGGCT	TTCTCATTAA	TGATGCTAAT
216601	TATTTTGGTT	TGTACATGTT	AACCAGGTAT	TGTACAAAAA	TATTTCTTTT	GGGAATCCAT
216661	AATGGATGTA	TGGCTTGAAT	ACAAATAATA	CTGTCTCTTG	TAAGTGCATT	GGAAATTTTT
216721	CCCTGCCACA	TGATTTTCATG	GAAGGTTGTT	TCGTGTATGT	ATGACTGCAA	ACCTGACTAT
216781	TCAGATCTTC	CGCAACAAGA	CAACTTATGT	GTGCATTAAG	AAGTTGCTGC	CTAAAATACA
216841	TAACACTGTA	ATCATTGGAG	ACTTTAAAGT	AATTAATCAG	CTATGCAATG	CCACGCTCCT
216901	GTTATCTCCA	GAGGGCTCTG	ACATTGACAA	ATGGTGGCTT	TCTATTTGAG	ACGTAATATC
216961	TAAAAAGCTT	TAACAGGTTT	GTAGAAGGAT	TGAAAGAAAG	AATGGGAACA	TTTAGGTCCT
217021	TATGGTAGAA	TAAGCATTA	TTGATTAGTG	TGTAGAAGGG	AGAGGCATGC	CACTTCAGAG

Figure 8 (Page 67 of 73)

SUBSTITUTE SHEET (RULE 26)

83/162

217081 GAAACTTCCT TCCCCAGTA AACAAATCTA CCTAAAACT AATTTTATCC CTTCTTCCCA
217141 GGTAGCACTG GCTGTGTCTG CTGTCTCCTA TGGTTCACAG TGATTTATGA TGACCCCATG
217201 CATCACCCGT GCATAAGTGT TAGGGAAAAG GAGCACATCC TGTCTCACT GGTCTAACAG
217261 GTACAGTGCA CACCTTGTAC CTGTGGCCCA TGCAGAGGTC TCTAGGGCAG GGTGTGGATC
217321 TCCTCTGAGA GGCACCATCT TGGCTGCTCT AATACTCATG CTGATTAGAT CTTTCTTTTC
217381 AGCCCAGTTC TCCTGGACGA GCTGTCCCCA TAAAGGCGAT GGTCACATGC CTACCACTTT
217441 GGGCCATTTT CCTGGGTTTT TTCAGCCATG TCTGGTTGTG CACCATCATC CTAACATACC
217501 TACCAACGTA TATCAGTACT CTGCTCCATG TTAACATCAG AGATGTGAGT TTACTTACCTA
217561 TACTTCTACG AAAATGATAA TGGTAATAAG GAGAAACAGT TCTGTGTTAC CTATTACATT
217621 CTGGCTTTAC ATATAACCAT TAATTTAACC TTCACAATGA CCTTGAGAGA GGCATTGTTA
217681 TAATTCCTTT TTCACAGATG TGGAAACAGG ACACTTAGAG GTGAGATAAC TTGCCCCAGG
217741 TTGCACAATA CTAAGTGATA GAGCTGCTGC AGCATCCATA TTCTTAACCA CTATGCTATA
217801 CTACCACACC AGCTGATTCC AAAGCTTCTT TTAGAAATAA TATTGCTGGG CCAGGCATGG
217861 TGGCTCATGC CTGTAATTCC AGCACTTTGG GAGGCCGAGG CAGGCAGATC ATGAGGTCAG
217921 GAATGCAAGA CCAGCCTGAC CAATATGGTT TACTAAATAT CATCTACTAA AAATACAAAA
217981 ATTAGCCAGG TGTGGTGGCA GGCACCTGTA ATCCCAGCTA TTCAGGAGGC TGAGACAGGA
218041 GAATCGCTTG AACCCAGGAG GTGGAGGTTG CATTGAGCCA AGATCATGCC ACTGCATCC
218101 AGCCTGGGCG ACAGAGTAAG ACTCCGTTTC AAAAACAAAA AACCCAAGAA ATTAATATTG
218161 CTTTTATCTG GAGCCCAGAG TGATGCAGCT TCTGGCCCTC TTATCTGAGA CAGTGTCTCT
218221 TTAGTGTGAA AAAGGATGCT AATTTTCCCC CAAACAACCC ACAGTATCAT GGGGGTAAGT
218281 TAATGGCTGG TCTGTGTAAC TGACAAATTT TGGTGCTAAC GTATCTCTAT AACTACTCTG
218341 TATAAACTTC CTTCTTCCAG AGTGGAGTTC TGTCTCCTC GCCTTTTATT GCTGCTGCAA
218401 GCTGTACAAT TTTAGGAGGT CAGCTGGCAG ATTTCTTTT GTCCAGGAAT CTTCTCAGAT
218461 TGATCACTGT GCGAAAGCTC TTTTCATCTC TTGGTAAGGA TAAGCGTGTG GGCCCATTTA
218521 ACCAATCCCT TTTCTGCACA TGGTCTCAGA GGGTTCCCTG ACAGCATGCT CTCATTGCC
218581 AGGGCTCCTC CTTCCATCAA TATGTGCTGT GGCCCTGCCC TTTGTGGCCT CCAGTTACGT
218641 GATAACCATT ATTTTGCTGA TACTTATTCC TGGGACCAGT AACCTATGTG ACTCAGGGTT
218701 TATCATCAAC ACCTTAGATA TCGCCCCCAG GTAAGAGCTC TACCTGTTTT TTCCCCTCCT
218761 CCAGACCCCT CCAGAGGTGT TAGACCTCAG TGGTCGCCGT GAAACTCTTT AATGTTACTG
218821 ACATTGCACT AATGGCAGAA TGACAAATAA CTACAAATAT CTGTCTGTGG CCATTTTTAG
218881 AACACAAAT GTGGCATT TT TAGAACAACA ATTTCCAATC TTGGCCAGTA ATCATTTTGA
218941 CAAAAACCTT CCAAGCTTC CCTAACAGAG ATTGAAGTGT GTATGCTGGG AAAAGGCCCA
219001 CACACAGGTG ATTTGGAAAA GTTCCATGG TGTGTTTCAT ATTAGCTACC ATATATATAT
219061 ATATATATAT ATATATATAT ATACAGTCAC AATAAGCCAG CTCCTGTGCC AAGACTTGCC
219121 ATATATCAAC ACATCTAATC CTCACGTTA TATTAGGTAG GCCCTATTGT TATCCCCATT
219181 TTATAAGGGA GAAGGCTGAG GCACAAGGAG GTTAAATGGT GTGACTATGG TCACATAAAG
219241 GCAGAGCCAG GATTTGGACT GGGGGAGTCT GGCTTTGGAG TCTGTGTCCT GCCCGTTGCA
219301 CAACTGGCT TCTCCACTGA GCAGCCGGGG TAAAGAAACG TGGTTCCAG AGAGACTGCA
219361 TTGCTCCCTG GTTATTGACT TGGTAGATTG GTAATTTCAG GTTTGGCAA TAGACATTGC
219421 CCTGAATGTC TTTAGGTGAA TGAAAACTG CATTAGCAA AATGACTTTG CCATTAGAGC
219481 TGAATTGCAT TAAAGTTGAG TTGCTGCAGA AGCTGTAGGT GGCTTTCTAT ATAAAATCAT
219541 TTATAAAATC ATCTTCCAC AGATATGCAA GTTCTCAT GGGAACTCA AGGGGATTG
219601 GGCTCATCGC AGGAATCATC TCTTCCATG CCACTGGATT CCTCATCAGT CAGGTTGGGC
219661 CAGTTTATTG AACATCTTCA AGTGGCAGGT ATTGTTTTAG GTGTTGGAGA TACACACGGT
219721 GCTCTAAAGA TCTGGATGGC AACACAATTA CTCTATTTAC ATGAGCCTCT AAATCAGACT
219781 CTGGTAGGTC AGATTTCCCA GAGGAAGAAA AATATAAGCT TATTTTCTCA AGATGAATAG
219841 ATGTTAGATT GATTAAAATG AGCTGTTCCG GTGCAGAAGA CAGCACGTGT GACTTCCTAG
219901 AGGTACATGA GCATGAAACA GTTCTTAGTT ATGACCAGAA TGAAAGACAC ATGTCAAGGA
219961 ATAGCAAGAG ACGAAGACAG AGGGGCAAAA GAAGATCATG AAGAATATGT TCAGACTAAT
220021 CCAATTTTAA AAAATCACA AAAGGGAAC AAAGTGTCTT AGGCCAGTTT AAAGATAATT
220081 TAATGCTGG AAACAGATCG GCTGTGAGC ATTGCAAGGA GGCTTGCTCG GTGTTGGAA
220141 ATGCAGGCTC ATGAGGAAGA TGAAAGACA GACCCAGGCA GGGATGGAAG GACTGACGAG
220201 AACCAACTTA CAAAGAGAAG TTTTGT TTTT ACTACATTTC TATGTGATCA AGTTCCAGG
220261 TTAATATTTG ACTAACTGC TAGGAATCCA CTGTGACTAT AATGCTGGAA ATGACTTAGT

Figure 8 (Page 68 of 73)

84/162

220321	AGGGCTTTCT	GAGGAGGGTC	ACACAGAAGA	CCAAAGAGAA	CTCATGTTGA	ATTGAGATGG
220381	GTTGTAGTGA	TAGTTGTCAA	CAGCCAATAC	AGAAACAAAA	AAAAACAAAA	CAAACAGCAA
220441	CAACAACAAC	AAAAAAAAAC	AGAGAAGACA	CAAACACAAT	GCCACAATGC	CATTTTAGGC
220501	ATAATTTTAA	ATGAGTAATA	TTATATGTTG	AAATCCAAAT	TTTCAGAAAA	ACATTAGTGT
220561	ATTTTATTTT	TGTTTAAAGA	AATAACCATC	TCAACTCAGA	ACCCCATGTG	CATTTTGGCC
220621	ATTTTGTTC	CAATAGTTTC	ATAAACTTTC	TTAAGTAACT	ACTGCACATT	GTTCTTATA
220681	TTCCTTGTGA	TCAACATTGC	AATACACAAC	TGGGAGGGCT	ACTAGAAGTG	GTGTAGAAGG
220741	AACCTGTGAG	ATTGATCATT	TTCTCTGTTT	TTTACATCTA	GGATTTTGAG	TCTGGTTGGA
220801	GGAATGTCTT	TTTCTGTCT	GCTGCAGTCA	ACATGTTTGG	CCTGGTCTTT	TACCTCACGT
220861	TTGGACAAGC	AGAACTTCAA	GACTGGGCCA	AAGAGAGGAC	CCTTACCCGC	CTCTGAGGAC
220921	ATAAAGTTAC	AAACTTAAAT	GTGGTACTGA	GCATGAACTT	TTTAAACATT	TTTTACTTCT
220981	CTCCATATTC	CTGACCATAG	ACTCAGCAGT	TCTTAACTCT	GGCTGTGTGT	TAGTCTTCCC
221041	TGGGGAGCCT	TTATAAGACA	CTGATACTTG	GGACCCACTC	CAGAGATTCT	GAATGAATTG
221101	GTCTGGGGTG	GAACCCAGAT	ACTACTAATT	TTTAGATACT	CCTTAGAGGT	TTCTAGCATG
221161	CGCCCGGGGT	TGACAACAGC	TGGACAAACT	TGAAAAGTCA	ATTCATGTGG	CCTTTGAATT
221221	TTCCTCATTG	GAAAGTACTA	AATAAATAAA	AATTCATGTG	AAAATGATCA	CTGATAAATA
221281	TCTTCATGGT	GGGGCAGGTT	ATTGGATGCA	GAGAAGATCT	GCTCGGAATT	GTAGCCATAT
221341	GTTACAGATC	TCAGCACCGA	TCGGAACGTG	AAAGCTATAA	TCCCCAGAAT	TAAAGTTTTT
221401	ATTATTTTTT	ATACATTGTA	AAACATAGAC	GTTTATTTAT	GTGATTAAAT	TCTATTAATA
221461	TTTACATGCT	AAAATAAAAT	AGACCATTTT	CAAATTATTT	AGATCCAGAT	ATTTCCATCA
221521	GATTAAACAG	ATATTTATTT	ATCCTAGCCC	AATTGCAAGA	GATTAAATGAT	GAGAAAATGA
221581	CCAATACAAG	ATTAAATAAA	TGAGGTTAAC	TTAGAAATCA	AGGACAGAGA	AGATAGAACT
221641	GGAAGGCTTG	TATTGTGAGA	AGAATGAATG	TGAAGGAAGG	CAATGTAGAC	ACTTCCAGAA
221701	GGGATAGCAA	TATAGTTTAG	ACCATATAAT	GAAAATTGGA	GAGAGATGAC	AGAGACACTT
221761	TCAAGTGAAA	TGACAATTTA	TATGGGGGAG	AAAAATATTG	AAGACATAAC	AAGATGAGAA
221821	AAGGCATAGA	AATGTATCAC	ATACAAGGCA	TAGAAGTGTA	TCACATACAA	GAGAAGTTCC
221881	TTTTGAGCGT	AGAAAAAGAT	AATTTAACCT	TCTTCATATT	TTTCTTACTT	TCCCAAGATA
221941	CTCAGATAGG	CAGCGTCAAC	TCTAACAGGA	ATTAATTTGG	CTCCTAACAC	TTAAGACATA
222001	TCCTTTAGTT	TGTCTCCTCA	CACAGAACTG	ATTCTGGTTT	TGCCACAACA	TGTCTAGAGA
222061	AGAAGTTCCC	ACCATATTTT	AAATCCTATT	AAAAAECTGC	TTGGACAAGA	ACCTTGGGTT
222121	AATTCAGCAG	ATGAAGAGAA	TCTCCTAATG	CAAATCAATG	GGTATTTTTG	AGCAAGTTTT
222181	TCAGAAAAAC	AGAGTGTGAG	GCCCTGAGGG	TGGTACTAAG	ATGAGAACAT	TGATTTTGCC
222241	TTCATGATAT	TGACAACACA	AAGAGGAAAG	GGGGTTTGCA	GAAAACTAAA	AGAAGAAGTA
222301	GAAGAAAAAA	GAAAGACATA	GATATAATAGG	TAGTCAAATT	ATGTACAGAA	AAAAGAGAAA
222361	AAAAAAACAA	AAAAGGGTGG	GGGACAGACA	ACCCAACTAA	AAAATGGGCC	AATGACTTGA
222421	ACAGGGACTT	CATAAAAGAG	AAAATGTAAG	TGGCTCCTTA	ACATATAAAA	AGATGTTCAA
222481	CTTCATTAGT	CATTACAGAA	ATGAAAATCA	AAACTACAAT	GAAATACCAC	TATAAAATTA
222541	ACTAATGGAT	AAAATGAAAAG	GAGATGGAAA	ACAAAATGTT	GCCAGACATG	TGGAGCAACT
222601	GGAACTTTCA	TACGTTACGA	ATGTGAACTT	TGGAAAGCTG	CTCGGCAATA	TCTCCTAAAG
222661	CTAAATGTAC	AATCCAGTG	ACTCAAACAT	TTTACTTAGA	AATGCACATA	TACATCCATA
222721	AAACATGTAC	AACAATGTTT	ATAGGAGCAC	TATCTGTAAT	AGCCTGAACA	GGAAGTTGTC
222781	TGTTAAAAAA	AGAAATGAGTA	AATAAACCCAC	GGTCTATTTG	TATAGCAATG	AGAATTAACA
222841	GACCCCAATA	TATAATAGAT	GAATGGGTCT	CATAAGCACA	ATATTGATTA	AAGGAAGACA
222901	AAACGCACAT	TCTTTTAAAG	GTTTATAAAA	TACTTTTAA	AAACAGCTAC	AACCAATCTG
222961	TCCTGTATAA	AATCAGTGAG	CGATTTCCTT	TGTGCAGGGA	TGGGGGTTGT	GGCTGGATGG
223021	ATGGTACTTA	AGAAGTGCTC	CTGGGGTACT	AGAAATATTT	TATTTCTTGA	CTTGGATGTG
223081	TGTTTACTTT	GTGAATATTG	TACATTTATG	ATTGTGTCAC	GTTTATGAAT	GTAGAAAAATA
223141	AAACAGAAAG	CAAATTCAAA	GTATCATCCT	TTTGAGAGCT	TCTGCTCTGA	CTTCGTTTTG
223201	ACCAATGGAG	CAGTTGGGAA	GGGGTCTTGG	TCCTTCGGTC	CTTTGCTTTT	TTTTTTTTTT
223261	TTTTTTTTTT	TAGACAGAGT	CTTACTCTGT	CGCCCGGGCT	GGAGTGCAGT	GGCTCGATCT
223321	TAGCTCACTG	AAAGCTTTGC	CTCCCGGGTT	CATGCCATTG	TCCTGCCTCA	CGCTCCCCAG
223381	TAGCTGGGAC	TACAGGCACC	TGCCACCATG	CCCGGCTAAT	TTTTTGATT	TTTTAGTAGA
223441	GACGGGGTTT	CACCATGTTA	GCCAGGATGG	TCTCGATCTC	CTGACCTCGT	GATCCGCCCA
223501	CCTGAGCCTC	CCAAAGTGCT	GGGATTACAG	GTGTGAGCCA	CCGCGCCCGG	CCCCTGGTCC

Figure 8 (Page 69 of 73)

SUBSTITUTE SHEET (RULE 26)

85/162

223561 TCTGCTTTCA TGTTCTTCTT GGTCTGTTC CTCCTCCTCT TTTGTTGGAA CTTCCAGTAT
223621 CAGAGCAGGA AGGAAGGCAA TGGGTCAATC GATGCTGTCA GCTTTTGGAT CAAACTGCAA
223681 GTTCTCAAAC AGCAAAATTA ATGAGCTCAG GCTTTGAAGA AACCATGACC CTGAAAGCAT
223741 CAGTTGCTTC CAATTGCATC AGTTGCCACG GGTGATAAGA ACAATGATGA CTCAGAATGC
223801 CTAGGTTTTC CCAGCAGCTT CTCTGAGGTT TTCCAGCAG CTTCTCTGAT TGATTCTCTGA
223861 CAGATGACTT CGGTGTGTCA GACTTTCAGG GTATCTTTCC TTATGTGATG GTTTGAGGAA
223921 GAGTTACCAT TCACATTCCT AATGGCTTCA GAATAGATGC AATTGTGAAC TGATAGGAAA
223981 CATTTCTAAT TCATCTCCCC TCCCCATCCC TAAAGGATTG TTTCTAACAA TAGTCATGAA
224041 AATTAATCA CTTTTCTCAA ATAGTTTATT GTCATCTACC TAATGATGAG ATGACTTACT
224101 TTTTCTCCTT GACTGTTAAA TATTATGAAT TATATTAATG TATTTCTTAA TGTTGAGCTT
224161 TCCCTTGAAT ATTCTTTTGA TGTACGACAG AATTGATTG ACTAATAGTT TATTTAGGAC
224221 TTTGGCTGAT GTACTGATAT ATGAGATTGG CTCTGTATGC ATACATGTGT TTTGTGTATC
224281 TTTTTTGTGT CTGGATATGG AGCTTATGCT GATTTCAAAA ACAAGAAAGG AGAAGTTTCC
224341 TTTTCCCCA TTAATCTGAA AAAGATTGAC TAGAATGGAA TTTTATAAT TGCTGTTGTT
224401 ATTTGAAAGC TTGAAAGCAT TGGTTGTAA AAATCATGCA GGCTGAAAGC CATTTTGAGG
224461 AGACTTTGAT AACTTTCTCA ATTTCTTCA GTTACTGGTC TTTTAAGGGG TTTTATATTT
224521 TTCTTTGATC AATTTTGACC ATTTATGTTA TCTTGGAGGA TCATCTATTT TACACACTAT
224581 TTAAAGTATA TTTGCAAAAA TTCAACTGTT TTATCAGGCT ATCTTTTAA TAATATATTC
224641 ATTTTATCTA TATCTGAGGT TTTAGCTTCT TTGTACTTCT GACCCAATTG CATGTGTGCT
224701 TTCTTTCTCC TTCATTAGAC TACTTAGTCA TTTACTAAT TTAAGAATAG CTGTCTTTT
224761 ATTTATTTAC TTATTTATTT TTGAGACGGA GTCTCACTCT GTCAACCAGG TCTGAGTGCA
224821 GTGGCGCGAT CTCGGCTCAC TGCAACCTCC GCCTCCCGGG TTCAAGTGAT TCCGCTGCTT
224881 CAGACTCCCG AGTAGCTGGG ATTACAGTCA TGCACCACCA TGTCTGGCTA ATTTCTGTAT
224941 TTTTAATAGA GATGGGGTTT TGCTATGTTG GCCAAGCTGG TCTCAAACCTC CTGACCTTAG
225001 ATGATCTACC CACCTTGGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCGCCC
225061 AGCCCTGCTT GTCTTTTAT TTTATATTTG ATTAGCTTTA TCTTTTATCA AGCTTATGTC
225121 CTATTTCCCT TTGCTTTACT TCATATAAAT TTTGTTTTGG ATAGTTTATT TATTTTTCAT
225181 TTAATTATGA AACAGGTTAA AGCTTAGAGG AAAATTGCTC CTCTAAGTCC AATTTTGTGG
225241 GCAGATTACA TTTTGCTGTG TTGTGCTCCC AAATTCATTG TTCTTTTAAAT GCTTTATTTT
225301 TCAAGTTAAT AACCTATATA GTAAAAAAGT GGCTGTTGAC TCTCAGCTTT TTTTTTTTTT
225361 TTTTTTTTTT GTAGATACAG GGATCTTGCT GTGTTGCTCA GGCTGGTCTG AAACCTGCTGG
225421 CTTCAAGGGA TCCTCCTGCC TTGGTCTCAC AAAATGCTGG GATGACAGAC ATGAGACACC
225481 ATGCCTAGCC ATGTCTCTCT CCTTATATAT AATAAGAAAA CAGACACACT GAGGCATCCT
225541 ATCATCTCAC TCTTGGTTTC ACTACTGTTT TCTGGAAGTT TTGCTCTGAC CTTTGTGAGT
225601 TAATGTATTA ATTTTGCAAT GAGTAGTTTC CATAGAAGAA TTATAGCATT TGCATTCTGT
225661 TGGGTATTAT ACTTTTCACT GTTATTTGAA CATAATTTGA GGGCTGAAAC CAAGATGAGG
225721 CAAGTGAGGT GCCCAGGAAG CAATATTTAA GGAGGCATCC TTTCTTAGGC TCATGCAAGA
225781 ACAGAATTGG CACATGAGAG TGAGTGCCTC CTTAATTTTG AGTGCTGGAC ACTTCTTGCT
225841 CACTTAGCAT ACCCCTGGAC AATGAAGTGT TTTTGTGTTT GTTTTTTCAT GTCCATCCTT
225901 TATCCTTCTT CATCTCAAAA CATTTCATG GAGTATTTT TTGGAGCAGT ACTTGGATGA
225961 GCCTCTGAGT CCCACAGTAG CTGAGAATTT ATTTCATAGT ACTCTTTATG ATCACTGTGG
226021 AGCCTTAAAA CATTGTAATA TTAACCTAGC TGGGAACAGA AATTTTGTTC CACAATTTGT
226081 CTTATTCAGA ACAGTATTGA CTTCCTGCTA GTCTCTCTG ATGTCCAATA TGAGGAAGTC
226141 TAGTTAGCCA GCTACTTTTT GTAGGAGAGC TATGTTTAGG CTAGGTGCTA TAGGATTCTC
226201 TTTATCCTGG AATTCCTTCA CCAAGATGTG CCAAGGTGTT AATCATTTTC TCTTGCTTTT
226261 TGGCTGGTGG TCTTAGAGTT TCCTTCGATT TTGTTTTATT TAGTGATTGT CCTCAATTTG
226321 TTTTCTTTAC TAAGAATCTC TCTTCTATTT ATCTGTATGG TAAAACTTG TTGCCCATCT
226381 TTCTGGTTTC TGCTGACTTT CATTTTTGA CTTTCTACTT TGCTTTCTCC ATGGACTTTT
226441 TGGTAGTGGA GGCAGGCAAA CACTTTCCAA AGTCTTTCTC AATTTCCATC AATTTCAACT
226501 TATTTCTTAA AATTGCCTCA GAATGTGCTT ATGTCCACAA TATCCCTCCT TCCACTTTAG
226561 AAAGGAAAGG CATCCACACT TTATTTAGGT GCAATGCCTG AAGTGTAAC ACTTTCTGGT
226621 TGTCACAAA GGAGTACTTC CAAATATTGG TTTGGGGATA ACCTGCTAAT GATTAACACA
226681 TTCACCTTGG CTCTTGGTTT GCCTGCTCCC TCTTCTTTTA TCTGCTGTGT GTATTTTTTT
226741 TAATCACTGA GAATATGCAC AGTATTGTAT GTTTTATTAT AAGAGAGGAC TGGCCAGAGT

Figure 8 (Page 70 of 73)

SUBSTITUTE SHEET (RULE 26)

86/162

226801 GGAATGTTT TGAATTCAGA ATAAGTGAAG CAGTACAGGA TAGGAAGTCA TTCTTTTCAAA
 226861 TGAAGCTGGC ATATTTTCCC AGAGCACCAA ATTTCAATAT ATATTTAAAA AACTTGATAT
 226921 GAATGATACA ATAAAGTGGT TAGAACTTTT ATTAATAATA ACTTATGTCA TGAAATACTT
 226981 ATTCTAATTA TAGTCACTCT TCATCTTATT TCATCTTATA ACATGTTTAA TGTTTTCTTT
 227041 TATTTACAAA ACAATTTTATT TTTTGATGAA AAGTTTTAGA AATCAAGTTA AAAATATTCA
 227101 AAGGAATGCC TAAAGTTTTT AAAATTCTTT TACATGTTGT ACAATCAAAA GAGTCTGAAG
 227161 ACCATTTAGC TATCCAAATT GTTTATTTTT AAGCAGTATC CCTTCTAATA TTTACTATTT
 227221 ATAATCCTTA AAAATTTGCC TTAGCACAGG AGAATTGCTT GAACCCAGGA GACGGAGGTT
 227281 GCAGTGAGCC AACACAGTGC CACTGCCCTC CAGCCTCGGC GACAGAGTGA GACTCTGTCT
 227341 CAAAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA
 227401 CGCCTTAACA TTATTTGTTC ATTAATAACT TTCTTTAATA CTACTAGTTT CCCTTTCCCTC
 227461 TCAGCCCATT GTCATATTTT GATTTTTATC ACTTGCTTTG TAGGACATAT GAGGTTTTTG
 227521 TTTTTTTTTT TTTTGGGAGA TGCAGTCTCC CTCTGTTGCC CGTGCTGGAG TGCAATGGCG
 227581 CAATCTTGGC TCACTGCAAC CTCTGCCCTC TGGGTTCAAG CAATCTCCTT GCCTCAGCCT
 227641 TCCAAGTAGC TGGGATTACA GGCACCCACT ACCACGCCCT GCTAATTTTT GTATTTCTGG
 227701 TAGAGACGGG GTTTCACCAT GTTGGCCAGG CTGGTCTCGA ACTCCTGACC TCAAGTGATC
 227761 CACAATCCTT GGCCTCCCAA AGTGCTATGA TTACAAGCAT GAGCCACCTG CCCAGCCAGA
 227821 ATATATGTTT ATTTTGAGTC CTTTAACAAA GTCATAAGAA TTTTAGGAAT TCAGTTACTT
 227881 TCTTGAGAAA ATCTCTGAAA AGATGCCAAT AATTTGTAGC CAATTATATT GATTTCTCTT
 227941 TTTCATATTG AGAATTGTTT TTTAAAAAGT TTGTATGTGT GAAGATTTTT GCAGTGTAGT
 228001 TAAAGAAACC ACCTGTGTGT TGGTTAAGCC ATAAGTACAT GTATTCAAAT AAATTGAGGT
 228061 GGGGTTACTC TGAGAATCAA AGGAAAACCT GAAGAAACAG GCAGCCTCAA AAGGCTCTAG
 228121 CTGTAGCAAC TTGCTCCATT GTTGAAATAA ATAGGCTTGA ACTTGATTTT TCCCTCTACT
 228181 CAACATTTAA GGTCTCAGAA GATAATATAA TTGGTGAAAT TTAAGTAAAG TGCTCACTCT
 228241 TTTGCTTTAA CAAACCCTAG AGAGCTGGTA GGCAGAGCCT CAACAGACCG TTTTAGCTTC
 228301 CAAAGGGAGT TCAGGACACC ATGATTCACG ACCACAATAC ATCACACATA ATTGAGAAAA
 228361 GATAGTTCCA CCAAATAAAG TTGAAATGCT GACAAGAAGG GGTAAGAAAT CTTGGAATAA
 228421 AGTTTATATA AAATTTATTT TTTCTTTTTT TATTGTTATG GAATAGGACC AGTTCTACTT
 228481 AAGCCACCCA TTTGCCAAAA TAAAGTGAGA ATCGTTTCTT TTGGGGACTC CTCTTTGTAG
 228541 CTCCAAGTGC CACTAACAAT TCTTAGGACC TGAGCTATAA GCCAGGTGAT TTCAGTTAAT
 228601 ATGATCAATT ATTTCAATTA AATGGCTCTA ATGTGCAGAG GGAACGGAGC CCATCAGCAT
 228661 TCCCTGCAGG GAACTGCAGT GGCTTTTATC AACTTGAACA GCTAGCTTTC AACTGTTTTG
 228721 AAATCACTTT CAGGGTGGTC ATGTAGTTGC TTTTTTGAAG TCAGAAGATG ATTCTGCCTC
 228781 TTTTAATATG TGACTCCTCA GATTCAGAAA GTGCTCGCTA GTCTTAAGAG TGAATTACCC
 228841 TCAGTGGTCC AGCGCTTATG AATCCACATC TAACCCTATC CCCTGGGGGA ACTATCAGAG
 228901 AAATTGGTGC CATGGACATA AGAGGAAGGC ACAGTGAAGC AGAGAGCCCC GCATGATGAA
 228961 AATCAGTGA CAGCATCATT ATTTACAATC TTGTAATCAC CCAGGAGCAT GAAAATCCAG
 229021 GCCAATCTGG CACCATGAGC TCTAATTTTT GTTGGAGTTC TTGGAACCGA TTCTGTGAA
 229081 TGACTGTTTA GCCATTTTAG AGTGTGGCAT ACGTGGCTGC TGGCATAACAG AGGTTGGATG
 229141 TAAACGGGCC TTTGCCCTCT CTTATGAACA TAGACAGGAA CTAAACTGTG TCACATAGGT
 229201 TCCAAATGGT GGCCTGAATA CTATTTACAA CTAAGGTACA ATGAAATTGA GTAAGTCTTT
 229261 TCCTCTTTTG CAGATACCAT CATTATTCAT ATATTTCTTC AAAGTTAACT ATTTGTATTT
 229321 GGTAATTTTT AATAGAAATG TAATAATTGC TTCTCAAGTT TAGTCTTTAG TCTTAAGGTT
 229381 GATGCTCTCC ATGTCCTTCC AAAAAAAGGT ATGTTGCTTT TATTATATCC TCGCCTTCAG
 229441 ATGGGATTAT TCCATTTTGT TCTTTGTTAA TATATACTTT GAGCCACTTT TTTTGTGGCT
 229501 CTGGGTGAGA TGCTATAGGT ACAATGACAA GTGATACGTG TGTGTCCCTT GTCACAAAAG
 229561 TGGATAGCCT AAGTGGTGAC TTTTACCTCC ACTCCAAATA TATGTATCAC ACACCAGCCG
 229621 TATGCCAGGC ACCACTCTAG GTGCTAGGGA TACAGCAGTA AACAGACAAA TGCAACCCCT
 229681 GCCCATGTGA AAGAGAATAA GACAATAAAT AAGTAAAGTG CATGTTATAT GGAGGTGGCA
 229741 AATGCTAAAA AGAAAAATTA AGCAGGCAAG AGGACTCATT GAAAAGATGA CATTTGGGTA
 229801 AAAGCCCATG TATATATGTT CTATTGGTTT TATTTCTCTG GAGAGCCCTG ACTAATACAC
 229861 AATGACTTTG AGAAGTTACT GGCTTTTGAT TTATCACACT ATTCGGAGTG CTGAGAGCCT
 229921 TCTTAGTGTG TATTCAGTGT TTTAAGAGAG CTTGTGGATG AATAATAAAT AGGACAAAAT
 229981 TTATCCAAAC TTAAGCCTTG CTTTAGGTAA AAGGGCTCCT CTTACAAGGT AGAAGGTTAT

Figure 8 (Page 71 of 73)

SUBSTITUTE SHEET (RULE 26)

87/162

230041 TATTTGGCAT TTAAATCCAA CTGAAGACTA ATAAGACTAA TTAATTAATA GTTTTTAAAT
 230101 CACAACTGGG TGCAAAATAA ATGGAAGTGC CATGCTCGCC AAGTGTGCAT GAGTGGTGTG
 230161 CATGGGAGAC AGCACGAAGC TAATCCCACT CATCTTGACG GTTGCTCCAT TTTTCTCCTA
 230221 AAATCAGTAA GACAGAAGCT GGTCAGATTA TCAAGAGCCC TAGTTAAACA CAGCAGTAGC
 230281 ATTTGGAAGG GGTTGCTCTC ATTAGGCAGT GCCTGACCAC AACAAGAGAT GAACAAGCCC
 230341 TGTATCTGAA GCCATCATGC CTAGTTATGG TCCCCACTG TTCATGATGC CTGAAAGGGA
 230401 GGCCCCCTGC ACCCTAGAAA GCTGGGTGGG TTCTACTGTC TGCTTTACTG CTAAAAACCC
 230461 TCTTCTTTGG ATCTGGACTT TACCCTATC TGATTTTTTT TTCTAATATA TGATTTGGCA
 230521 CTGAGTCTGT CACTGCTGCT AACTCAGCAG TTCTAGGGTC ATTGCCCCAT TGCCTCACAG
 230581 AAAGAATTTC ATAGCTTCCA GCATCCTCTC TCCTTCATTA TACTTTGATT TCAGCATTGC
 230641 TATTTTTTCT CTTGGGTGTT GCAGCTCTCT CTCTCCTTCC CATGTCTTGT TGGTTTTCTG
 230701 CTAACCTCTG CTTTTTTTCT TTTTTTTTTT TTGAGACGGA GTCTCGTTCT GTCACCCAGG
 230761 CTGGAGTGCA GTGGACAAT CTCGGCTCAC TGCAACCTCC GCCTCCCGGG TTCAAGCTAT
 230821 TCTCCTGCCT CAGCCTCCCA AGTAGCTGGG ACTACAGGCG CTCACCACTA TGCCCCCTA
 230881 ATTTTGTAT TTTAGTATT GCTGTCATCA ATCCACATGT CCAGAAGCAC CTAGAACTC
 230941 TAATTCCTTG TAGGTATCAA ACCCTAGGAC TCTTTCCTCT AATCACAATA TATAATCCCT
 231001 GATTCCCAAA CACGGTCTTT TCATATACAT TTTCCACTGT ACATACTTTC TGACCTGGAA
 231061 AGCTCTTACA CAAACAGGCC CTCCCCTAGG AAGCCTTTAT AAATGTTCCC AGGAAGAATC
 231121 AGTCACCCAA CAGTGTCTTT GTCACATCTT AGGTTCTACA CCTTTATTTG TTCTATCTGA
 231181 ATGTAATCTC CCAGAGGGTG TTATCATCTT TTTTTTTGAG ATGGAATCTT GCTTTGCTGC
 231241 CCAGGCTGGA GTGCAGTGGC ATGATCTCGG CTCACAGCAA CCTCCACCTC CTGGGTTCAA
 231301 GTGATTCTCC TGCCTCAGCC TCCTGAGTAG CTGGGATTAC AGACGTGTGT CACCACCT
 231361 GGCTAATTTT TGTATTTTTA GTAGAGACAG GGTTCACCG TGTGGCAAG GCTTTCCTCG
 231421 AACTCCCAAA CTCAGGTGAT CCACCCGCCT CAGCCTCCCA AAGTGTGGG ATTACAGGTG
 231481 TGAGCCACCA TGTCCAGCCC CATCTTTTTC TTTTAGTTTA GTTCTTAACA AATAGTCTGA
 231541 CACAAAGTGG ATATAACAAT ATTTTGAATT ATGAATAACT AAATGAATAT TTCCAGATTT
 231601 CCTGGTGCTC TCAAAGTTT ATGTTACAAA AGAAAAACAA GTCTAAAATA CCTGCCTCAA
 231661 GTTTTTATCT GTACTATGAT TTCAAACCAA ATAAAAACA GGTGGGGTAA AAAGTGAAC
 231721 AGGAAATACA TATACTGAA AAATTTGGT ATGTTAGTAT GATAATACTA GGTATTATTT
 231781 CCTGTTTCCC CAACTTCATT TTCTATAGCA ATAAAAAGAA ACAAGTAAAT GTATATTAAT
 231841 TTAATTTAAA AGAAGTAGTC TACCATCTCT TCTGTTAAAA AGAAAAAGT ATTTTAAAAA
 231901 ATTATCTCTG GAAGGATACA CAGGGAACAT TGCTCTGGTT TCTTCCAAGA GAGAAATGAG
 231961 GAACTAGAGA GCATGGCCAA GTGGGGTTTT GCTTTTGTTT TTGTTTGTCT ATCTGTTAGC
 232021 TTTTATTAT TTTCTTTTGT AGGTTTGAAT TTCAAACCAC ATAAATCTGT TACATGCTCA
 232081 TAATAATAAG TTTAAAATAA AACTTTTGGC TGGGTGCAAT GACTTACACC TGAATCCCA
 232141 GCGCTTTGGG AAGCAGAGGT GGGAGGATAC TTGAGGCCAG GAATTTGAGA TCAGCCTGGG
 232201 CAACATAGTG AGACCCTGCC TCTGTAGAAA TAAACAAAAA TTAGCTGGAT ATGGTGGTGC
 232261 ATGCTTGATC TCCTAGCTAC TTGGGAGGTT GAGGCAGGAG GATCCTTTGA GTCCAGGAGT
 232321 TTGAGGCTGC AGTGAGCTAT AATCACCCAC TGCACTATAG CATGGGCAAT AAGGTGAGAA
 232381 CTTGTCTCAA AAAAAAAAAA AGGGGGGGGG AAACAAATAA ATAAATATAA ACAAACCTTT
 232441 TGTTTCAAAA TATGTAATAT TTAGCACTAA AGAATTCTGA ATTGTAGAGC TAAAAAGTAC
 232501 TTAAAAGTTA ATAATTATTG TCTCCTTTAA AAGAATTGTT ATCAAAGTAT AATTTTATC
 232561 CAGAAAAATCA TCCATATCAG CAAGCTAAAC TTTCTCAAAA TGACATATCC ATGTAATTAG
 232621 CTCCCAGGTA ATTAGCAGGC AGCCTCTACT CAGGTTGAGT ATTCTAATC TAAAAATTGG
 232681 AAATTCAAAA TGCTCCAAA TCGGCAACTT TTTGAATGCT AACATGATT CCAAAGGAGT
 232741 GCTCATGGAA TATTTAGAT TTTGGATTTT TGGATTTGAG ATACTCAGTA TAATGCAAC
 232801 ATTCCAAATC TGAAAAAATC TGAAATACTT CTGGTTCTAA GCATAAGGGA TACTCAACGT
 232861 GTGTTAGCTA ATTAGACCCT TCATGGTCTC TTCTAGACCT CAGCTTCTTC AAGGTAACCT
 232921 CTATCCTCAC TTCTAATAGC ATGAACTTT CTGTTTTAGA ATAATTGGA TTTTCAGGAA
 232981 AGTTGCAAG ATAGTACAAA GACAGTACAG GAGAGTTCCC ATATATCTTT CACCTAGCTT
 233041 TCCCCATTG TTAGGATTTT ACATTATTAT GATACATTG TCAAATATAA GCAACTCACA
 233101 TTGATACATG AAACCTTATT AACCAACCC TAGACTTTAT GTGGATTTCA CCACTGTTTC
 233161 CACTAATGTT TTCTTTCTGT TCCAAGTCC AATCTGGAAT ACCACACTGC ATTTTCTTGT
 233221 CATATCTCCC TAGTCTTTT TTGTCTGTGA CAATGTCTCA GTCTTTCTT GCTTTTCATG

Figure 8 (Page 72 of 73)

SUBSTITUTE SHEET (RULE 26)

88/162

```
233281 ACCTTAACAG TCCTGAAGAT CATTGCTTT TTTTTCATAA TTACACCGGA GTTATAGATT
233341 TTTTGAAATA ATACCACAAG GGCAAAGGGC CCTTCTTGTC ACATCATTTT AGGGAGAACA
233401 TGATATCCAC ATGACATCAC TGATATTAAC CTTTCATCATG TGGTTTAGGT AATGTTTCAG
233461 GTTTCTCTAC TGCAAAGTGA TTTTTTCCC TTAATTTAGC CCACCTGAAC TTATCAATTT
233521 TGTTTTCTTC CATGACTAAT ACTTTTGTTA TTATAGCTAA AACTTCATTG GGGCCAAATC
233581 TTAGATCATG TAAATTTTCT TCTATATTTT ATTCTAAAAG CTTGTAATGT TTGATACATT
233641 CTAAAAGATG TAATGTTTGA TACATTACAT CTAGTCCTTT GATTTATTTT TAGTTACTTT
233701 TGTATAAGGT GTGAGAGATG TCTCCAGTTT CACTTTATTA ACACATTGTG GTGTTCCAGT
233761 ACTATTTGTT GCTAAGACTA TCTTTTTTCC ATTGATTACC TTTGCCTTAG TTGGCAATAT
233821 TTTTGTGGT TTATTTCTAG ACTGTTTATC TCATTCCACT GATTTGTGTC TATCTTTTGTG
233881 ACAAACCTGT TGATTACAGT AAGCTTTGAA ATAGTTCATT TTTTGTGTCA ACTTGACTGA
233941 GTCAGGGGAT AACCAGCTAT CTGGTTAAAC ATTATTTCTG GCTGTGTTTG TGAGCGTGGT
234001 TCTGGATGAG ATTAGCCTTT GAATAGGTGA TCCTAGTAAA GTAAACTGTC TTTCCCAGTG
234061 TGGATGGCAT TATGCCACCT GATATTACAG GTCTGAATAG AAGAAAAGGC AGAGGAAGGG
234121 GGAATTTGGG CCTTTTTTTC TGCCTCACTG CTTGAGCTGG GACATCTCAT CTGGTCTCCT
234181 GCTCTTGAAC TGGGATTTAC ATCATCAGTT CCTCTGGTTC TCAGGCCTTC AGATTGAGAC
234241 TGAATCATA CACCAGCTTT CCTGGGTCTC CAGCTTGCAG ATTACAGATC ATGGGACTCC
234301 TCATCTTCCA TAAATGCATG AGCCAATTCA GTCTATGTCC TTGAAAAC TGCCCCACTGCA
234361 GATTAAGGCT TTTTTCCTACT AGGTGAAATA AAGAAGCTTG TTAGACAGAT TTCCCTTCAT
234421 CCAAGTCCCT CTCCTCTTTA AGTTACAACA CATTGGCTAC ACCTAAGTGC AGGGGTGGGG
234481 ATGAGGTAT AGTCCTCTTG TTTGCTGAGA AGAGAACTGT ATTGGGAAAG CTCTAGAAGT
234541 GTTTGATACA TACATAAACA AGGCATGGTT TTTGCACTTA ATTTACATT ACATTTTTC
234601 CAGAAAAAAA GGAATGTATA GGCATCACGT AACTGTACTA GCTGGAGTCA TTCTTCCTGA
234661 TTATCAAAGG TAAACAGTTA TTAATCCTAT ACCAAGATGT CAAGGAGAAG TACTTTTGGGA
234721 ACACAAGGAA TTCTCTGGGA GTCCTTACTA CTCTCAAGCC CAGTGAAAAA GTTAATGAAA
234781 AACTATAGTA CCTTCCTATA AGCTGGATGA CTAATTACCA GGCTCATTTA GGAATTTGCC
234841 TTACCAAGTA AAACATAAGG GCAGCTGAGG TGCTGACTGA AGACAAATGG AGCATAGAAT
234901 AAGAGTAGTA AAGAATGCCA AAAATGCTGT CATGTATCCA TTGACAAAAG GAGCTATAAA
234961 GCCTTTAGGT ATTTTCACAC TTGCTCTGTT ACGTAAATGT ATGTGTGTGT GTGTGTGTGT
235021 GTGTGTGTGT GTG
```

Figure 8 (Page 73 of 73)

SUBSTITUTE SHEET (RULE 26)

89/162

```

1  CACACACACA CACACACACA CACACACACA CACAAATGAG GTATATAAAG GGTCTCCTAA
61 AATGTCATCT GATATTTGTT ATTTTCATATT CTCAGATTTT TAATCCATTT AGGTAGGTCT
121 ATTTTAGATA GCCTTGCTCTG AAACAGAGCT GGGACCTGAT GAGTGAAAAT GAGCTCACCA
181 GAAGAAAAAT CAAACAGGCA TTTCAGAGAT TGAGGCCAAG AAGTTAAATG TCTTAAATGG
241 GCAGAGCTTA GCTGCTTGAT GTGAAAAGAG ACCAGCGTGG CTGGAACAGC AAAGGAGAAC
301 AGCAGAAGAG GTGAACAGAG GCCAGAGATG GTCACGTAGT GGGCCCTTAA GTCATGGTAA
361 GGAGTATGGA GAATGAATTA TTGCATGTAT TGAATATGTA GGTGACGTGA CTCACAGATA
421 CTTTGGATTT GTAGAGATGA AGGAAATGTA GCAAGTGACA CTCTTAGAAT GTTGATTGTA
481 GTAAATGGTA GTGTGAGTTA TTGAACTGGG GAGAACTGGA AGGGATAACA GGCTTAAGGA
541 GCACGTTTAT TCCTGTGTCT TGGAACTGTT TAGGGTGAAA GACCTATTAG AGTTCCTAAAT
601 GGAGATGTCA AGTGAAAATG TGGCTACACA CATTTGCATT TCAGAAAAAA GGTGAGGCTG
661 GAGATGTAAA ATTGGAAGTT TACTGCATAT AGATAGTCTT TGGAACCGTA GTATTGATGA
721 AGCCATTAAT GAGACAGAAC AAAGACTAGG GACCAGAGCC AAGCTCCAAG TTTCTAAAAAT
781 TTAGAGGATA GTATAGTCTG GTCATTTTGA GGTGAATACT TAATAACAGA ACAATTTGCT
841 GAAGTGTAAG TTTAGAGCCC TACACTTTTA GCTCTGACTA TTAACGAATA CAGGAAAAGAA
901 TGGATATGGT TATCTGCCTG GTGTCTGTGA AATAATTTAA GCCAGGAAGA GATCCTCACC
961 AGAAACTGAC TATGCTGGCA ACTTGGATCT TAGATTTCCA GCCTGCAGAA TTGTTAGAAA
1021 ATAAATGTCT ATCGTTTAAAG CCACAGTCT GTAGTATTTT GTTATGGCAG TCCAAGCTGA
1081 CTAAGTTTTG GTACCCAGGC GTGGGATGCT GCAACAACAA ATACCTAAAC ATGGGGAAAGT
1141 GGCTTTGGAA ATTGGTGATG GGTAAAGGCT GGAAGAGTTT GAGGTTCTA CTAGAAAAAG
1201 CCAATTGTGA AGGGACTATT GAAAGAAATA TGGACATTAA AGGCAATTCT GGCAAAGGCT
1261 CAGAAAGGAA GAGAGCTGGA CAGAAAGCTT CCATTTTCAT AGAAACTTAG ATTTATAACG
1321 ATCATGGATA GAATATTAAA TATGCTGGTT AAAATATGGA CTTTAGGCCA GGCGTGGTGG
1381 CTCACGCCTG TAATCTCAGC ACTTTGGGAG GCTGAGGGCA CAGATCACGA GGTGCGGAGT
1441 TTGAGACCAG CCTGGCCAAT ATGGCGAAAC CCTGTCTCTA CTAAAAATAC AAAAATTAGC
1501 TGGGCATGGT GATGTGCTTC TGTGGTCCCA GCTACTCGGG AGGCTGAGGC TGAAGAATCG
1561 CTTAAACCCG GGGGGTGGAG GTTGCAGTA CCCAAGATCA CACCACTGCA CTCCAGCCTG
1621 GGATACAGAG CAGGACTCCA CTCCCCCGC CACACACACA CAAAAAATAT ATATATATGG
1681 ACATTAAAGT CAACTCTTGT GAGGTCTCAG ATGAAAATGA GGGACAGGTT ATTGGAACCT
1741 GTAGAAATCA CTGTTCTTGT TACAATGTGT CAAGAACTTG GCTGAATTAC GCTGTAGTGT
1801 TTAAGTGGAA GAACTTATAA GCAGTAAAC TGGATATTTA CCAGAAGAGA TGTCTAAGCA
1861 AAGTATTGAA GGTGTGATT AGGTCTCTCT TACTGCTTAA AGTGAAATGT GAGAGGAAAG
1921 AGCCGAAATA AAGAAGGAAT TTTTAAAGCA AACACAATCA GAACTTGGAG ATTTGGGATA
1981 GATTTCTCAA TCTATATTGT AAAAATTGAG AAAGTTTTTC TTGAAGAGGT ATGGTTGAAC
2041 AATGTTTTCT TTTTCTTTT TTTTCTTGGT TTTATTTTFA TTTTATGTT TTTTGAGACA
2101 GGGTCTGGCT ATGTCATCCA GGCTGGAGTG CAGTGGCACA ATCTCAGTTC AGTGCAACCT
2161 TTGCCTTCAG GCTCAAGCAA TCCTCCACC TCAGCTCCTT AAGTAGCTGG GACTACATGT
2221 ATGCACCACC ACACCTGGC TAATTTTTTG TTGTTGTTTA TAGAGATGGG GTTTTGACAT
2281 GTTGCTTAGG CTGGTCTCTA ACTCCTGAGC TCAAGTGATC TGCCCTCCTC AGTCTCCCA
2341 AGTGTTGGGA TTACAGGCGT GAAACACTGA GCCTAGCCTG AACAACCATT TGATAAAGAG
2401 ATAATGGGTG TGACCCAAGG ATTTAATCAG CCATCTCAGC AGAAGCCAGG AAGAGAGATG
2461 GGATTATTCC AGCAGAGACA CTGCCAATTT AAACCTAACGT AGGCAGAGAA AACAGAAAGG
2521 AACAAAGGAA GGTTGTCGAC TTTTGAATT CTATAGAACA GGATCATAGA GCTACCTGGC
2581 TGTCATGTG TACTATTCTT TAAGAAAAGG AAAGACTGAC CCACCAAGG CAACCTACAA
2641 GATCACTAGG GCTGACTCTT TTTTGTTTT TCTTGAGGCA GTCTCACTGT CACCCAGGCT
2701 GTAGGGCAAT GGTGTGATCT CAGCTCACTG CAATCTCCAC CTCCAGGTT CAAGGGATTG
2761 TCTTGCCTTA GACTCCCAAG TAGCTGGGAT TACAGGCTCT AAATCTGTAC CCTCCGAGT
2821 AGCGCTCCTG CCACCACTTG CCCAGCTAAT TTTTGTATTT TTAGTAGAGA TGGGGTTTCA
2881 CTATGTTGGC CAGGCTAGTT TGGAACTCCT GACCTCCAGT GATCCATTCT CATTGGCCTC
2941 CCAAAGTGCT GGGATTACAG GCAGGAGCCG CCAGGGCTGC CACTTTGATG TCAGACTCAG
3001 AGAGTACAGA TGGGATAGGG TGGGGGTGGG AACATGTAGT CAAGGCTGAC TCTACCTGTT
3061 TCAAAGATGC CCTGCAGAAC TGTGTGGGAG TCTCTCACAG ATGGCTGCCT GGGTGGGACC
3121 CCACCAAACT GAAAGACCGA GACTTCAGGC AGGGCAGATG GAGTAGGCCA ACTACAGAGC
3181 CAGAGGTGAC ACTGAGACAC CACTGGCCCT GGAAATCAGG GCATCAAGCC AAAGAGGGTT

```

Figure 9 (Page 1 of 74)

SUBSTITUTE SHEET (RULE 26)

90/162

3241 TTTCTTAAGA CCTAACAGAA TTTGCCTTGC CAGGTTTTGG ACTTGATTAG GACACATTAC
3301 ACCTTCCTTC TTTCTATTT CTCCATTTTC TAATGGGAAT GTCTATTATG CCTGTTTCAC
3361 CATTGTACCT TAGAAGCATG TAACATTTCT GGTTCACAC GTTCAAAGCT GGAAAGGAAT
3421 TTTGTCTCTG GATGAATCAC ACATTGAGCC TCACCCGTAA CCTGATTTAG ATGATTTTTT
3481 AGATGACACT TTGAACTTTA GAATTGATGC TAGAATGAGT TAAGACTTTC AGGGGGCTGT
3541 TGGGATGGAA TAATTTTTTT TTTTTTTTTG AGACGGAGTC TAGCTCTGTC GCCCAGGCTG
3601 GAGTGCAGTG GCACCATCTT GGCTCACTGC AAGCTCTGCC TCCCGGGTTT ATGCCATTCT
3661 CATGTCTCAG CCTCCAGAGT AGCTGGGACT ACAGGCGCCC GCCACCACGC CTGGCTAATT
3721 TTTTTTTTAT TTTAGTAGAG ATGGGGTTTC ACCGTGTTAG CCAGAACGGT CTCGATCTCT
3781 TGACCTTCTG ATCCGCCTGC CTGCGCTTCC CAAAGTGCTG GGATTACAGG TGTGAGCCAC
3841 CATGCCCCGGC TGGGATGGAA TAAATTTATC TTGTATGGGA GAAGGACATA CATTTTGGCA
3901 GGTCAAGGAC AGAATGTTAT GGAATAAAT GTGTCCCCCA AAATTCATTT ATTAAAACCC
3961 TAAACCCAG TGTGACTGCA TTTGGACATA GAGCCTTTAG GGGGTACATA AAATAAAGA
4021 TCACAGGATA GGGCCCTAAT CCCATTGGGG CTGGTGTCTT TACAGAAGAT GAGACACTTA
4081 GAGCTCTCTC TCCACGCAGG CACCAAGGAA ACACCATACA AACACACAGT GAGATGGCAG
4141 CCATCTGTTA GCCAGGAACA GATTCTCACC ATAAACTATG TTGGCACCTT GATCTTAAAC
4201 TTCCAGGCTC CAAAACGTG AGAAAATGAA TTTCTGTTCC AAGCCTCTTA GATATGGAAA
4261 AAAAGATTCT GTTGTTTAAG CCATCCAGTC TCTGGTATTT TGTATGGCA GCCTGAGTAG
4321 GCTAAGACAA TGAAGGATGT GGTAAACTT TACGTCCCAA CCACATACCA AAGAGGCTGG
4381 AATTTAGCAT GCTTCTTCT TCAACTGTA GGCAATGTGC ACAAGTTCTA AATCTAAGA
4441 CATGTTGGCT CCTTACTCT GCCCAAATA CAACTCAAAC AAACAACGTG AATATAATAA
4501 CATCCAATGA AGTTCTGACA TTTCTTCAAC ATGAGTACAG TAATTCAATG CCAGAGAATT
4561 CATTTTATTT TGAAATCTAC ATGCCATATT CCAATTTCTG TTGAAGATGC AATGGTTATA
4621 TTTATTCTTT TTAATATAGA TTTATCAGAG TGGGCGCGGT GGCTCATACC TGTAATCCTA
4681 GCATTTGAGA GGCTGAGGTG GGCAATACAC CTGAGGTCAG GAGTTTGAGA CCAGGCTGGC
4741 CAACATGGTG AAACCCTGTC TCTACTATAA ATATAAAAAT TAGCTGGGTG TGGTGGTGCA
4801 TGCCTGTAGT CCCAGTTACT AGGGAGGCTG AGGTAGAATT GCTTGAACCT GGGAGCAGGA
4861 GGTTGCAATG AGTGGAAATC GCACCAGTAC ACTCCAGCCT GGATGACAGA GCAAAATAAT
4921 AAATACATAA AATAGATTTA TCAGTTTATC AATAATATAG TTTTCTTTTC TAGGTGTAAA
4981 TATAGGTAAT GACTGTCCTT TAGTACATTT TCTCATGATG CTCCTCTTAC TTGGTTTGGT
5041 ACAATATTAA GTATTGAAAT AAAATAGAGA ATCCTGTGCG TACACATGAG CACTTATTCC
5101 ATTTGCTCAT CTCCAATATG CACGGGAAAT TCTCAAATTG CTAATAATCT TGTAACACAC
5161 ATGCATTATA TTCAACAGGA ATATATAAAT TTATAATTAT AATTTAGGAT CAACAGATGA
5221 CAAACCTTTA GAAGGTTTGT ATTTAACCTT AAAATATAAT TTTTAAAAA TTGGTTATAA
5281 AATTTCTAAT ACTTTCTTTT TTGTGACCTC AAGGGGAAAA TATAATCTT ATAAAAGTTC
5341 AAATGATTTA CAGAATACAA AAAGTGAATA GAGATGATGA ATGAATTAAA GGAAAGGATA
5401 TTGCTACATA GATTTGGAAA TTTAAAAAGG GAAATTACGA TTGTTGATT TGTGTTAAAC
5461 TGATCTGCTT TGTTCAAGAT ACCTTATGTA CCAAAAAATG ATTTTATCTC AGCCTCATAT
5521 CTCAGTAAAT TCCTGAGACA AACTTTAGTC CCTGGTGCCC AGGTGCCCTT GGTAATTGGG
5581 AGACCTCTAG GTTTAGCATC CTCATCCACT CGCCCCAATT TAAATAGTCC TCCCCAGGGC
5641 CATTGAGGCA AGGGAGATGA AAACCTGCTC AAGAGTTGGA ATCCAATTGA AGCTACCGAA
5701 ATTCATTGCT CAATAGATAA TTTTCCTGG AAGTAACTAG GGCTTTTGAA TATAATAGTG
5761 GGCATTTCAA AGTAGAAGGT AAAGTATTTT GGAGATGAGG AGACAGGACA GAGCTACGAG
5821 GAATGTCTTT TGCTCAGGGA CTAGGCTCTT AGCAGTACCT CTTAGGTAAG AACTGGTTAA
5881 CTGGCACCTT CTGTGTTTCT CTGAAGCTCC CTTTGCTTAG GGACTAGGCT CTTAGCAGTA
5941 CCTCTTAGGT AAGAACTGGT TAACTGACAC CTTCTATGTG TCTGAAGCTC CCAGAACAAA
6001 CTGCCAATGA AATTTGGATT TTTGGAATAT AGTTTCTTTT TTGTTGTTAC TTTTGTGTTT
6061 GTTGTTTTTT TTTGAGAGTC TCACTCTCAC TGCAACCTCC CCCTCCTATA TTCAAGTGAT
6121 TCTCTTGCCT CAGCCTCCCG AGTAGCTGGG ACTACAGGCG TGCACTAGCA TGCCAGCTA
6181 ATTTTGTAT TTTTATAGT AGATGGGGT GGTTTTTTTT TGAGACAGAG TTCACTTTG
6241 TCGCCCAGGC TGGAGTGCAG TGGCAGCAT TTGGCTCACT ACAACCTCCA CCTCCCGGGG
6301 TTCAAGTGAT TCTTCTGCCT CAGTCTCCTG AGTAGCTGGG ACTACAGGCG CCTACAGGTG
6361 AACACCGCCA CACCTGACTA ATTTGTGTAG TTTTATTAGA GATGGGGTTT CGCCATGTTG
6421 GCCAGGCTGG TCTCAAATC CTGACCTCAG GTGATCTACC CACCTCAGCC TCCCCAAGTG

Figure 9 (Page 2 of 74)

SUBSTITUTE SHEET (RULE 26)

91/162

6481 CTGGGATTAC AGATGTGAGA CACCAGATCA GCCTCAGAAG ACATTTTCTA TTGGAAAGAG
6541 AAAACACTAT TAGCAACCTA TTAGTCTAAT ATTTAATACT TAATGTCTTC CTTAGTAATA
6601 AACCAACTCT CTACAACAAA GTGCTTCCTG GCTGCCTAGT CATTGATTCA TTCAGTTCAA
6661 CATTTTCTCA ATGCCCAACA GCCAAGTGTC TCCTGTATGC CAAGTTCTAT GCTGATTATC
6721 AGTATTTGAA TAAGAGGGGG TCTACATCTT AAGTACTGCT TAAGATGAAA GCCTCTAGGT
6781 TAACAACTT AACACAATGT ATCATTCACT ACTAAATAGA CCGAATACAA AATCTTGTTA
6841 TTGGAGCCCA GAGAGAAGAA TTGAAATTCA AGTTTTCTCT CTCTCCTTTT CTCCTCACC
6901 ACAATAAGTC AGTTGCACCA AGTCTGTAG CTCTTTACTG AGCCATGTTT TCACGTGTCC
6961 CTTTGTTTTA TTTGCCACAC CCTAAATAAA AATTGTACTG GCTTTTTTTC CCTGGGTTTA
7021 CAGTATTAAT ACATTGTCAA GATTACCTC TTCGTGTAGA TTCCCTGGGG AAAATTACCT
7081 TTCCTCCTTC CCTTAAATTC TTCAGAGGTT AGAAAGCCAT TAGTAACATT CTGGTATGTG
7141 GACAAAGTTT ACCCATTATG TATGGATGTT TTAGTCTTTC CATTTTCTG ACAATAATCT
7201 CTTAAGGAGG TGTGGTTATA GAATAGTCAG CTGTTATAAG TACTGTTTTT CTGGCCTTAC
7261 AACTTAAATT CTTTAAAGCTG TTTCTTAGTT TGCTCATCTC AAAATTCGGA ATAAGGATAA
7321 AACCTATCTC TTAGATTGTT GGATTAAATG AATTAACATA CTGGAAGCTC ATGAAATGTG
7381 CCTGGCACAC AGTAGTGCCT AATAAACCAT CTCTCTTATT CAGCCTGTTT TCTGATTTCA
7441 GAATCTACAC TTGCTGAGCC AGGTTCTTTT CATTTCAAGG TGAGCAAAAG CATAACAAGGA
7501 AGAGATGGAG GTAGGAAGAG ATTAAGCCCT AGGCCAAGGG AGCTGGAATC AAAGGCAATT
7561 TGGTCAGTGA ATAAAAAGGA TTCCAAGGCC CATAAGGCAA TTCTAACCTT AGGATCGAAA
7621 TTCTCGGACA TACAGGAAAT GCTGGGGGGG GGAAATCCG GTCTTCTCAG CCAAGAGCC
7681 ATGTGAAACC AGACCTTCAA ATCTGATGAT TCTCAGCCCA GCTGCCATT AGAATCGTTG
7741 TAATTTAAAA ATACCCTCGG AAAATTCTAA TATGTGGCTA TCAAAGGTGA TCATTTGCTT
7801 TTATGCCACT TTGTTTTTAC CCAAATGGGA CATCCAACCC TTTTCTTTG AGAGTAGTTG
7861 TAGGGAAAGG AGGGGGTGGA GGGAGGGAAG AGCGGAAAAG GCTGGATCCG CCCCAGCCG
7921 GTGTCAATG CTGGGAAGTG GGAGGCGCGT CAGCAGTAAA CAGCTTCTGC TAGGATTATT
7981 ATCTCCTGCC ACACACTCGG ATTTGAAGGC TCCAAACGAA ACAATGCAAA ACGCTTCAGT
8041 GGAGTTCCAG AAGCGTTAGA CTAACGACT GGGTCTGTTT GGCCAGTCTG AGCAGCTGGG
8101 CGCAGATGCA TAGGCAAGAC TTAGCCCGCC TAGACTTTTC TGCCCACTTA ATTCCGATCA
8161 AAGCAGAAAC CGGCCGGGCG CGGTGGCTCA CGCCTGTAAT CCCAGCACTT TGGTAGGCAG
8221 AGGCTGGCGG ATCACCCTGAG GTCAGGAGTT CGAGACCAGC CCGGCTAACC TGGTGAACCT
8281 CCGTTTCTAC TGGTGGCGGG CGCTTGAAT CCCATCTACT AGGGAGGCTG AGGCCGGAGA
8341 GTCGTCTGAA CCCGGGAGGC GGAGTTTGTA TGCAGTGAGC CGAGATCGCG CCACTGCATT
8401 CCAGCTTGGG CAACAGGAGC AAACTCCGT TTCAAAAAG CAAGCAAACA AACAAAAAA
8461 TGCAGAAACC GAGATCCGGA AGAAAACCTC GGCGAGATTC ACAGAATCCA GGAAATAGG
8521 TCTCTAGAAA TTTGTCCATG GTCCCAGATC TCCATTCTT GTGGGTGGGG CAGCTGTTAC
8581 CAGATCCCTA GAAGCAAAGG TTTTTTGGG GGACCGTGTC TCACTGTTGC CCAGGCTGGA
8641 GGGCAGTGGC ACGATCTCGG CTTACTACAA CCTCCGCCTC CCAGGCTCAA GCGACTCTCC
8701 TGCGTCAGCT TCAAGAGTAG CTGGGAGTAC AAGGTATGTG CCACCACGCC CACTTATTT
8761 TTTTATTTAT TATTTTATT TAGTAGAGAG GTGTTTACC ATGTTGGCCA GGTAGTGTC
8821 GAAGTCGTGA CCTCAGGTGA TCAGCCCCCT CGGCCTCCCA AAGTGGTAGG ATTAGAGGGG
8881 TGAGCAGAAA GCAAAGGTTT TTGAGTGCC ACAGGCCCCA CTCTATTTCC TTTTCTGCCT
8941 GTAATGGCAA CCTAGACGCT TGAGCTTCTT AAAATACAAG AGTAAGTTGC ATGTCAGGCA
9001 CCGTTCTACA TTAGGGACAT TAGTCTGTTT TACAGACACC TTCAACTCC CTGGTTAACT
9061 TTTAGGTAAT ATACTCTGCA CTTTAGCAGG AATGGAACCT ATAACCTCA CAGAATTAGG
9121 AAAGTGAGGC TGCCTACAGC CTAAATTGAG AAAAAATAG ACGGGGACT AGTCGGAGGA
9181 CCAAACAAGG TTACCAACAC GTTAGAGTTT TGCCTTCAAT TTACATTTT AAAGTAATCA
9241 CAACGAAGTG TTTAGATCAC GAGGCATCCC TGCATGTAAA CTGTTAGGCA CTAACATGG
9301 TCGATCTTAC AAAGCATTAA CTAGAATATT TCTTTAGAGT ATGATAGTAC GTAACGTACC
9361 TACTATTACA TACAAACAGA CCAACCTTTA GTAACAGCGC TCCCCAAAA CCGAAAAGCA
9421 GTAATACGCT TTGCTCAAGG TTGGCATAAA ATTAACCTAC CTTAGTGCCT TTTTCTCTC
9481 TACCTACAAG CAGTGAGGTT AGCTCTTCTT TTGAAACGGT AGGGGGGCTC TGAAAAGAGC
9541 CTTTGGGTTT GATAGCGTTT CCGGGAGCTC AGATACCTGT CAAATCACTT GCCCTTGGCC
9601 TTGTGGTGAC TCTCGGTCTT CTTAGGCAGA AGCACGGCCT GGATGTTAGG AAGGACGCCG
9661 CCCTGAGCAA TGGTCACCCG GCCTAGCAGT TTGTTGAGCT CCTCGTCGTT GCGGATGGCC

Figure 9 (Page 3 of 74)

SUBSTITUTE SHEET (RULE 26)

92/162

9721 AGCTGCAAGT GGCGCGGGAT GATGCGAGTC TTCTTGTTGT CGCGAGCCGC GTTGCCGGCC
9781 AGCTCCAGGA TCTCGGCGGT CAGATACTCT AACACCGCCG CCAGGTACAC CGGCGCGCCT
9841 GCCCCAACCC GCTCTGCGTA GTTGCTTTTA CGGAGCAGGC GGTGCACTCG GCCCACC
9901 AACTGGAGAC CAGCGCGAGA AGAGCGGGAT TTCGCTTTGG CGCGAGCTTT GCCTCCTTGC
9961 TTACCACGTC CAGACATTGC AATCAGACAA AAATCACC
10021 GAGAAAACAA ACAAATCAA GAAATATGTA AAACATGGCC GCTTTTATAG GTAGTTCCTG
10081 GGGAGTAAAT CCGACTTTTT GATTGGTCGG TAGCAAATGC TAGTCAGATA GCCAATAGAA
10141 AAGCTGTACT TTCATACCTC ATTTGCATAG CTCTGCCAC GGATGACAAC TGTGTAGTTT
10201 GTCTTCCAAT TAACTAAGAG GTACTCTCCA TCCCTCATTA GCATAAAAGC CCTATAAGTA
10261 GCAGAAATCC GCTCTTTACT TTCGACATC TTCTGGTGTT TTAAGATGCC TGAGCCAGCC
10321 AAGTCTGCTC CCGCCCCGAA GAAGGGCTCC AAGAAGGCAG TGACCAAAGC GCAGAAGAAA
10381 GATGGCAAGA AGCGCAAGCG CAGCCGCAAG GAGAGTTACT CTGTGTACGT GTACAAGGTG
10441 CTGAAACAGG TCCATCCCGA CACTGGCATC TCTTCCAAGG CCATGGGCAT CATGAATTCT
10501 TTCGTTAACG ACATATTTGA GCGCATCGCG GCGGAGGCTT CCCGCTGGC GCATTACAAC
10561 AAGCGCTCGA CCATCACCTC CAGGGAGATC CAGACGGCCG TCGCCTGCT GCTTCCCGGA
10621 GAGCTGGCCA AGCAGCCGT GTCCGAGGGC ACCAAGGCCG TCACCAAGTA CACCAGCTCC
10681 AAGTAAACAT TCCAAGTAAG CGTCTTAACA CCTAACCCCA AAGGCTCTTT TAAGAGCCAC
10741 CCAGATACCC ACTAAAAGAG CTGTGGCCAG ACGCCAAATT TTATTTGGCG GCGGAGGGGT
10801 ATTAGAATGT AGGAACTGGA GAGGGGTGGG GACAAGTGTT GCAGCTTAGA GAGGGACAAA
10861 GGGTCTGAA CCCGAAAGAA GCCAGCCATT AAAAATGGGT TTGGGGTCAA TTCGTTGTGC
10921 TTAAATTTAA AATGGGGACA AGCGGCCATT TTGCTAACTC GGCGTTCCCG GAAGAAACCG
10981 CAGGCTCGCT TAGGTTTCAG ACCCAGCTGT CTGTCCCTGT CTACGTCGCC AGGATCAACG
11041 GTTGCCGTAA TGTCATAATT TCGCCACCAG CTTCTAGCCA ATAGGCTGTC CTGTCATTTT
11101 AAATATTAAC CAATCGAGGG AAAGCTGTTT TGAGACTCTG ATTTACATAG CGGACCGGAG
11161 TGGGAACCTG GGCAGTAACT GCCTAAGGAA GGACTCCCC TCTGTTTTCG TGGCGCACAC
11221 CTTCTGTAGTA TACTGAAGGG TGTGTCTCCT GGGTTTCCAA CTGCCCCGGT AATAGTCTTT
11281 TAACCTAATA TCGTCACTT TTGATAACAA CACTAAGGCA GTACAGAACT AAAGATGTAA
11341 GCACTGCGCC AGATGTTGCT TCATACATCT TATTCTATTC AACTGGTTTA TTCAAGATTC
11401 AAATCAAATC AAATTTTGCT TGAATCCAG TGCTCAGTCA GCCATAAATG GTGTGTTGCC
11461 TGATTGAAAC TTAAATCTC CGTAGGGGGC TTGTAACATG CAGAAAAGTT TGAAAGTTGC
11521 TTTAGGAGAA GCCAACTCTT AACTGCTGGG TAAATTGACA AGCCTTCGAA CACTGAAGTG
11581 AAGGCCAGTA AGGACTAGGC GCTGGGTGGG GGAGAATGAA GAGGAGACGT CATTAAACTT
11641 AGCACATACA CTGTGTCTCC TAGAGGACTC TCCCTTCTTA GACAACTGCA GGCCGCTTTG
11701 TGGCTGGGA AATTCCACAT TCCCTTAAGT ATTTTACTCA TGGTCTTTTC CAGGTAAAGA
11761 TTTTAAGATG AAGGTTAGA CGTAGTCTAC CTATCTTTTT ATTCAAGTCT AGAACACGTT
11821 TTTAGCACCT AGAAGTTTGC TTTCTCCATT AAAAACC
11881 AGTGTTAAAG CAGATTTTTA CAACTTAAA TACCATGTAA TTTAGGTTAC AGTTACTTAA
11941 CATAAGGACT GTGTGATCTT AAATCTGCAA TTTCTTTCAC ACCTGGGAAA TAACTAAGG
12001 CCTGTCTTTG GTGCCAGACA AGGCCTTATA CTTGAACACT GCTGTGCAAT CACAGGCTGC
12061 CTTGCCTAGA TAACTTATCT GAGAAATTCT GATGAGAAAT GAAATTTCCA GAGTCCCTCA
12121 CAAGTAAATT TTTTTTCTT TTTTTTTTTT TTTGAGACGA AGTTTCTCTC TTGTTTCCCA
12181 GGCTGGAGTG CAATGGCGCG ATCTTGCTC ACAGCAACCT CCGCCTCCCG GGTTCAGCC
12241 ATTCTCCTGC CTCAGCCTCC GGAGTAGCTG GGATTACAGG CATGCGCCAC GACACCCTGG
12301 CTAATTTTGT ATTTTATAGTA GAGACGAGGT TTCTCCATGT CGGTGAGGCT GGTCTCGAAC
12361 TCCGGACATC AGGTGATCTG CCCGCCTTGG CCTCCCAAAG TCCTGGATTA CAGGCTTGAG
12421 CCACCGCGCC GGGCCTAAAT GGTTTTTTTT TTTTCTATGC CTCTAATGGA CCTGGTCACT
12481 TATTCCCATC CAGACTGACC GCTCTCTAC CTGCCAATA ACTAATCAGT GTAACCAAAA
12541 TCTGCAACA AAATTCAGTA TTCTTTCCCC GCCTTTTCCC CTTTCTCTTA CATAGATTAT
12601 GTTTTTGCCT GTGTTAGATG AAATAATTCT ATTGCTTGTT CTCTCTCTG TACAAGTACC
12661 CAGTAAGCAA ATTATTAAT TCTTGGTCAT TTATTTCTGA ATTTTCCACC AAGACAGTGT
12721 TTATGTGAGT CATACAATAA GAACCAACAG AAATGTGTGT CTTGGAAACA GGTGTCTAT
12781 CCCTGGACCC TTTGAGTTT CTGTTCACTT TCCTTTGGCT TTTGCATGCT AAAAGTTTAT
12841 CGTCCGCGTT TGTTTGTCT GGTATTCTA ATTGGACTTG GCTGATTGGT TGCATATTGG
12901 TGGCAGTAGT AGAATTGAA TTCTGGTTTT CTGGTCACAT CATTAGTGA TTAGTCAGT

Figure 9 (Page 4 of 74)

SUBSTITUTE SHEET (RULE 26)

93/162

12961 GAGAGGACAG GAAATCTGGT TTATTTATTA ACCTTTTTTT GGGGTGTTTT TGTGTTGAAGA
13021 TGTTGATATT CTCTGTGAGG ACACAGGGTT AGAGTTGGTG TTTTCTTTT TGACTTTTACA
13081 TGGGATTTGA TGTGTTGTGC TTGTATGCCT CTTTCCACCT TCCAAAACCT GTCTTTTTTG
13141 AGTCCAAATA GTTGTGATA TCTGCAAAAC CAGTATTCCT GTGTTAAGAT GATATGAATA
13201 TAAAATGGCT GCCCTGTTAT AACTTTTGAC TTAAAGAAAG TGTTAGGACT AACAGGAGAC
13261 AAAAAGGAAA TCAAGGAAAC CAAATGTCTG GTCTCAATAA CTGCTATGGC AGAGGCTCTA
13321 CAGCTTATTA TTAATTTTAG TAATTTTACA TTATTGCCCC TTCACGTTCT TTAAGTAAGG
13381 TTAGAGGACA GAAGAAACAT AATGTTGTTA CAAATTGGAC TATTGAGTCA GGAACAAAAA
13441 AGAGTGCTTT CAATATCTGA ATAAACAA GATTAAATAT TTTCTAAACC TTAACGAGTT
13501 TATTGTAAGG GATGTGATGC TGGAAACTAG GAAACTAGAA TTTTCTTCTA AACTGAGAAT
13561 CAGAATTATT CATATTCTCA GCAGTGGTGC CACCTGAGG ACTTCTGATC TTAATTACAT
13621 ACTTTTATTT CTTTAACTGA TCAACATGCT AAATAGATAA CCTATGGCTC TGTTTTTACC
13681 CACTTTAAAT TCTGTTCTAT TAGCACGGTT AGCTTTCCTA ATTGGCAATA AGATTGAGAC
13741 TATCTTTTTT TTTTTTTTGA GACAGAATTT TGCTCTGTGG CCCAGGCTGG GGTGCAGTGG
13801 CACAATCTCG GCTCACTGCA ACCTCTGCCT CCAGGGTTCT AGCAATTTTC CTGCCCTCAGC
13861 CTCCCCAGTA GCTGGGATTA CAGGTGCACC ACCACGCCCT GCTAATTTGT GCATTTTTAG
13921 TAGAGATGGG GTTTCGCCAT GTTGGCCAAA CTGGTCTCGA ACTCAGGTGA TCCACCTCGG
13981 CCTCCCAAAG TGATGAGATT ACAGGCGTGA GCCACCGTGC CCAGAAAAGA CTATCTTATT
14041 TTATGAATTT AAATAATTGT GAAATTATCC ACTTAAGGGA ATTAATAAAT TATAATGTAA
14101 TCTTAAATTT TAGTTGGCTT ACATAAAGAC TTAAATACA TCAATTTAAA TAAAACTCA
14161 TTTGTCTAAA AAAAAATCAA AAATTTTCTT TGTGCTTAA ATGTGCTACC TCTTTAAGTT
14221 CTAATTAAGA GAAAAAAGT TTAAGTGTGA GTTTCATTAG TGGTCTTAGT TAACAGCTTA
14281 AAGTATTTTG TAAAAAAT ACTTCACAAT TTTTAAATAA CTTAAAAATA TTAATACCTC
14341 TTTTATTAGG TTTTTTTAAT AAGGAAAATA TATAATACAT CTAATCAAGA TTATTTTTTG
14401 GACAAATTGG CTTAATAATT TCATTTTAAA AATGGCTTCT TTATCTTAT ACTGTAAAAA
14461 TAAATTATAG AGAATATTAT AGTATACACA AGTTTAGGGT TCATATTCTA AAAACAAAAA
14521 ACAAAGCTA ATTTAACTTG CATTTACTAA ATTTCTTCCA CTAGTTGTAC TGGTTACATG
14581 AGTTAACATC ACTTTATTTA TTATTTTAAA ATTGTAAAT ATTCAATTGAA CCAAATTAAA
14641 TGATAATAGA TAATGTCATT TTTAAAAATG GAATTAAAT TTATGTTACT AATTATAAGG
14701 ATTCAATGTG TGAGCTTAAG TACTGAGTTC ACAGTGTATG ATAACTTTAA GAATTTAGGT
14761 GAATATTATT AAATGAGTA AATTAATTCT CAATCTTTGG ATACCTGGAC AATTTCTAAA
14821 TTGGAGGGTA CAAATACAA ATCACAAGAA ACAGTGTAGT TTTATGCAAA TAACATTTTT
14881 ACACAGTTTA GAATAACCAT TGATAAACAG ATAAGAGAAC ATATGATTGC CTTAGAATAG
14941 ATACTGTTGC TTTCCGCACT TTAGATTTGT AAATCATGTA CTGTATACGT GTGGGCGTAG
15001 AGGACCATGC AGGTTTTGGA TGACTGCCTC TGTTTTCGTC ATGCCTATGC GGAACACAA
15061 TTGCCTGCTT TGTTTAAGGG CTATGGTTAA TCCAAACAGC TCTGACTCTA TCAAGTACTA
15121 TAGCTACAGA GAAACACAAG TAAGCATTCG AGATAATGAC TACCTTGAGC CTTTACTTAT
15181 TTAATAAGTT GTTACTGTTT GTTAATGTGG TACATTCAT TTACTATGGA TTGTCACCTC
15241 AAAATAAGAC TTCAATCTTT TTCTTATTTT TATATAGCCA TGATTTATAT TCATATCTTA
15301 ATGTAATAAC CAATCTTCTC TGACAACATT ATAACAATGC TGGAACCTCC ATTTTCAGTA
15361 CTTCAAACAA CAAATACTGC TTTTATACTT CAGAGCAGAT GGATATGTGC TTCCAGTGT
15421 AAACACATTT GGAATCTCAC TGAGAAATAC ACTATCACTA AAAATACAGT TCTGAGATTC
15481 ATTAAGAGC CTCCAGAATT CTGGAAGTAG GAAGTTTCCT CTTCAAAGTC TACAGAGGAA
15541 GACGAGTCT GAAATAGACA GCTTCTTCTT TCTTTTACCT GTGGTATTAT TCTGTTTTGT
15601 CCTTTTCTCC ATTATCTGTC TTTCCAGTGA TGAATTTTG ATCTGGCCCT CCCAAGTATT
15661 AAAAAACAAG CAAATAACA AATCTCAGTT ATATTTTACT AAGATATTGG CATGCTAACT
15721 TTTTGCAGGT TTGTAACAAG GACCTTTATA ACTTGACTAA AAGTTCCTAA ATAAGAATAT
15781 TTACTAGAAA ATTTATTTCT GCCTGTGGCC CACATTTGAG TCAAAATAAT CAATTAGGAA
15841 AAATGAACTT GTTTAACTAA AGTTGGCCAA ACTGATCTTT GAGACCTATT CATCTAAGAC
15901 AAGCCAATTA AATCTTGGA GACAAATTGT ACTTTAAGGA ATTCTTATA TATTTGTAAT
15961 TACCTCATA ACTTTTTTTT TGCCCTACTT CTGTGCTTCT CTAATATGCA GATTATTAAA
16021 TGTGTTACA AAGCCATTGT CAAAAAACA AAAACAAAAA AACTAAACAA ACTCACATGG
16081 TTAGACTTGC TCCTTTATGA GATATTTTAA CCAAAATGG AGGAGTTGAA AAACCTCTGGT
16141 GCCAGAAATC GTGAAGACAT GGCCTACCTA ACTTGGAAT GTTGGTTGTC AGTGGAAAAA

Figure 9 (Page 5 of 74)

SUBSTITUTE SHEET (RULE 26)

94/162

```

16201 ACTACACAGA GATAGCCATA GTGCTGCACA GCCAATCTTA AGTGTCTTCTA GAGAATCACT
16261 AATTGTTTCT AGAGAATCAC TAATTGTTTT CTTTAAACAT TCTTGGTTTA TACAAGAAGA
16321 GAGTATCCAT ACTAAACTCT TTTCTACTGA AAATAATGTG CAAACATAAC ATCCTATTCC
16381 TAGACAGTTT GTAGTTTTTT TCTCCCATT CTATTTTATA AATCATCTTT TTAATAACT
16441 TTGTTGAGTG AAATCAGTCC ATTGCTTGAT ATACCTTGAG CACAAGTAAA TAGTATGCCA
16501 AAAATTAAAT GTCTTTCAGT CACAGTTTGA CAAACTCAAC TACCCTGAGC CTATAGAGTG
16561 GTAATAATTG CCCTACTCAT AAAGATGGGG TGAAGATTAA ATGAAATAGC ACCTATAGAA
16621 CACTAGTTCC AGACGTGGTA TCATGCTAGT AAAATGGCTG CACAGCACTG CTCAATGATG
16681 ACAAAAAGTG AAGCTTCTGG AGACAGACTC CAAGTTTGAC TCCCAGATCA CCACATATAA
16741 GATGTGGGAC TCTGAGGCAG GTCATTTAAT CTCTCTGTGC ATTAGTATCC TTCTCTATAC
16801 CTTTACAGTG ATGGTAATAG CACCTACCTT CTAGAAGTAT GTGAAGATTA AAGATCCTTA
16861 ATGCATATAA ACCACTGTGT TTACTGCTGT TTGACAAATT TTATTTATAA CCATCTTTAC
16921 GCTCCTAAAA GGACTTGAAG CAGCTTATGA CTGAAGACTT TGGTAGGAGT TGGCCTTCTA
16981 TAAATTATAA GAATTCATA AATTATTTGA TATGAAAATG CCAGTTGATC ATAGTATGTT
17041 TACCGGGGTC CAACAGGTTG AGAAAAATA CACTTTTTTT CCCTGAACAT ATGAAATTAG
17101 CTCTCTAGGC ATATTCCTAA GGACTTAAAG AATGATAACT ATCATTCTCT TTAATCTTCT
17161 CAGATTTGGA AGGATATATA TATTGACAC ATTGACAGAC AATCCCAGTA GTCCTAAATT
17221 AAAAGACATT AAAAATTAGT GAACTTTTT CTACCTTTAG CCTGTGTAAT CCTGGATGAC
17281 CAAGCATAAA ATTAAATTGA GTAGAGTATA CCACTGTAAC ATTTCTGAA AGGTATTCTA
17341 GGCTCTGAGT AATTTCTTTG GGGTCTGAAG ATCAGTTTGA CATATCCTCA AGTATCATGA
17401 GTTCATTATA ATTAAGAAAA AGGGAGTAA TCTGGAGAAT GAGCCACTTT CTTACTACTC
17461 CTTGACCTCA GTTCTTTTTT TCAGAGACAG GGTCTCACTT TGTGCCCCAG GCTGCCAGGC
17521 TGGAGTGTAG TGGCGCAATC GCATCTCATT GTAACCTCCA CCTCTGGGC TGAAGCCATC
17581 CTCCTGCCCT AGCATCCTGA GTATCTGGAA CCACAGCAGG TGCACACCAC CATGCCAAGC
17641 TAATTTTTTA AAAAGTTTTT TGTAGAGAT GGGTCTTACT ATGTTGCCCA GGCTGGTCTC
17701 AAACCTCTGG GCTTAAGTGA TCCTCCTGCC TCAGCCTCCC AAATTGTTGG GATTACTAGT
17761 GTGAGTCACT GTACCCCGCC CCACTTCAGT TCTGAGGAGG AAAAAATAG TAATAATAAT
17821 GGGACTTTGG TTTGCTGATT TAAAGATTCA TGTAACCTTA TCATCCAATG CGCAATTTGT
17881 AGAATAATTA ATAGAGACAT CTGGTCTCAT GTTCTACAG TTGCTCATGC CTTGATAGTA
17941 GATCTCCTTG CTGCTGGCTC AGAAGGGTAA AAGAGCAGAA ATGATGGGGC TTCTCTCATT
18001 CTATGAGGAA ATAGACCTAT GTAGAGGAGG CTACCTGTGG TAAAACCTTA TCCTCATCAC
18061 TTAAATCTCT AGGCTTATTC TCTGACCATA TCAAGTTTTC AAATGGTAAA AGAATTGGAT
18121 TCAAGAGAAA TATGAATAAA CTTTTGTTTT CACTTTTCTC CCTCCTCTCC CCCATTCTC
18181 CTTCTCTTTA TTTTCTTGTC CTTAGTTTTT TTTTCACTTT TTTGTCTACT ATTATTTGCC
18241 CAAACTCAAC TGTAGGCTAG AACAAAAAAA AATTGAAAAT TAAAATGTGC CCCTTTTGT
18301 GTTAGACTTG CTTAAACAAT TGGGGTAAAT AACCTTGGAC ACTAGATTTT AAAACACACA
18361 CATTTGAGCT TCAGTGCACT GAAATAAATA TATTTTAAAC AATTAAAAAA TAAAATTGCA
18421 TGTTAAAAAA ATCTGCAGAG AACAATACAC GTTGTGAGAT CTTGAATGGA AGGAAAACTG
18481 CTAGCCTCAA GAGTGGATCA AAGATGCTCA GCAGGCAACA GAGTAAGAGC ATGTTGGAGG
18541 GTTTAGAGAG TGTGCTCAGG GTTCTAGGCT CTAAAAATCA GACAGTCCCC ACGGCCTGGC
18601 CTTGCTCGCT GTATCTTCTT TATGAAAAAC ACTAAGTCTT TTTCTCACT GGATAAATTT
18661 TTATCCTTCA AGTTTAGATC AAATGGAAT TTAGGACACT GACTAGGTTA CATTCTCTT
18721 TTAAGAGCGT ACAGACATTC AAGGGCTAGA GGATGTGGGT TTAGTGACA GGCTCATTAT
18781 CCAACAGCTG TGCTACCTGG GAACTTAAAC CTCTCTGTGC CTTAATTTCC TCATCTATAA
18841 CGCAGGGAGA ATGACAGTAG GTATCTCATA AGGTTGTTGG AACAACTAAA TGCATTGGTA
18901 TCTATTGTGT AAAGTGCTTA AAACACTGCC TGGCACAGAG CAAACATCCA TGAACTTTA
18961 GCCATCATCA TTATCATTGT TCTCAGAGTC AAATACAATA TCTCATATCT GATAAATTAC
19021 AGAAGTGAAT CAATCACTCT CTCTCTTTTC TCCAGGGGGA GACAACAGCT TTTAGACATA
19081 TCTTTTCCAA CAGTCGTCAC TGCTGGACAC TGTTTCATCT TGCAAATAAA CCAATGAAAA
19141 TGAGTGATCC TAGAAGAAGA TAAATGGAGG TATTTTGAAC AATCAAAGAA GGACAAATGA
19201 ACACCTGGCT GAGAAAAATT AGCTCTTTT TCTATGCATA AAACATATAA AATATTCTTC
19261 ATAGAAATTT ATGACACAGG AAACATAAAG ACAAATTAAT AATAACTCCT AGTATCTCCT
19321 ATTCTTTTTA TATGTATATT ATATATACT ATATTCTAT ATACATATAT CTCACATCAT
19381 GTATCATATA TAAAATAAAT TTAGGTGTCA TGATATATAT TTAGATAAAT ATACTTAGAA

```

Figure 9 (Page 6 of 74)

95/162

19441 ACTTTTTTAT GGATGTATAA TTTATGGATA TATTGATAAT TATGTATTTG TTATTGACTA
 19501 CTTCAATTGA TTCCCATTTT TATGCATTAT ATTATAGATT ATATAGCTCA CACATCTTTG
 19561 TACATAAATC TTTGTTCAAA TATTATTTCC TAAGGATAGA CTTTCATGAAG TGGAAATACT
 19621 AAATCAAAAG TGAAAAACAT TTTCTAAGGT TCTTAACATA TACATTGCCA AATTGCTATT
 19681 CAGGATCATA CCAATTTATA ATCCCAAAAT AATATGAAAA TTCCTGTTTT ATAGCACTCA
 19741 TATTTACAAT AAATTTTAAA AATCACTGTT AACCTAATAG TCCTTCAAAA GAAAAAAAAA
 19801 TTGAAATTAC ATTATTTTAA TGACTCTATT AGTGAGGGTC ATTCTTCCCA TGTTTTCTGT
 19861 TAGCCATGAC CCTATAAGAA ATAAACTGCA CTGCAAAATG ATAAACATGA TATCAATCAT
 19921 TACATGGGAA GGCACATAT AAAGAATAAT ACCTTAGGTT AAGGCCACAT AAATATTTAT
 19981 CAGGTGCCTT TTCTGCGGAG GACTCTGAAG GGATACTAAA CTGCATTAG CTGCATGCAA
 20041 CTGAAATTAC TTTTACCTAC ATTGTCTCTT ATAAACATTA TAACTACTCT TTGAGAAAAGT
 20101 GTTTACTATG GACTGAATTG TCTCCCCATC CCCCCAAATT CATATATTGA AGCCATAAAC
 20161 CCCAATATGA CTCTATTCCT AGACAGGACT TATAAGAGGT AATTAGGGT AAATGAGGTC
 20221 ATTAGGATGG GTTCCTAACT GGATAGGATT GGTGGCCTTA TAAGAAGAGG AAGATTCTGC
 20281 ACTTGGTCTT CCAAATTAAA TAATTTATTT AAAAGAAAAA AAAAAAAGA GGAAGAGAGG
 20341 GAGCTCTGCA CATATACTGA GGAAAGGCTA TGTGAGCTCT CACAGTGAGA AGGTAGCACT
 20401 CTACAAGCCA GCAAGAGAGC CCTCACCAGA ATCCAGCCAT GCTATACCTT GCTCTGAGAC
 20461 TTCCAGCCTC CAGAACTGTG ATAAATTTT GTTGTTTAAA CCACACAATC TATGGTATTT
 20521 TTTTATGGCA GCCCAAGCCA ACAAAGACAG CATCATTGCT GTCACCTACA GACAAGAAAA
 20581 CTAAGACTAG GAGAGAGAAA AGTTAAACTT GTCCAAGGTC ACAAAGCCA GAAACAAGTG
 20641 AGGTGAGAAG TTGACCTTGT TCTCTCAAT CCAAGGCCAG GACTCCTCCA CTCCACATGT
 20701 AGATAGCCAC CTCACAGTCA ACAGCCAAAT GTCCACACCC CAGAGTCAGC ATTAGACCAA
 20761 GATGTCTTAC CAGGAGACAA ATGCCTCATC TTGAATAAAT ATGTTCTAAC AACTTACCCA
 20821 TGTAACACAT TGAATCTCAT GAGAAACAAA AATGCAAAGT ATGTAGAAAA CTATGTTTAC
 20881 CACTTAACTG ACAGTGATAA AAAGCTTAAT GATATCCTTA TAGTCTTGGA GGGGTTTGTA
 20941 TATGTGGTGA AACAGGTGCT CACGCACTGC TGATAGACTG TAAATTGGTC CTAGAGAGAA
 21001 AAATAAATAA ACTGGAAGGA GTTATGCTGT ATGTTTACTT TTTTATGGA AACATATGAT
 21061 ATACCTGGAA ATTGCGATTG CCATGCATCT ATTTCTTCAA TGGGTATGCA CAGTTGAGCT
 21121 GTTCCCATGC ACCAGGCACT GTAATGGGAC AACTGCACAT GACAGTCAAA AATCTCAGTC
 21181 TCATGAAGTC GACATGCTCA TGGAGAGGTG CTACCCACTA AACTAATATT TGTATATCAA
 21241 TTATGGATAC ATTGGGCCAC ATTTACAGAA ATTCACTTAC AGTGGGTTAC CAGAAGGGAT
 21301 TTTTTTCTT GATTGGCAAG AAGGCTAGGC TGTTTTGTTG GGGGCTGGCA GGAGCTGCT
 21361 AGGCTGCCCA AGTATGCAGG TCTCTTCTAT CATCCTGTGT TAACCATCTT CCATGTATCT
 21421 TTCAACCTCA TGGTCATCTG CAGCATGTCT AGGGGTCATA TCTATGTTCC ATGCAGGAAA
 21481 AAAGGGTAAA GGGAAAGGGA AGTAGGCATG TACCATTFTA ATGCACACCT TGGTTTTTCA
 21541 AAAATTTAAG AAGAAAGACT TTCTGCTTTT CTCTGACTAT TCTGTATTCT GGATTACAAC
 21601 GCAACAGAAA CGTCACCTTA AATTCTAATG TTTTCTCTC CTGCTTTCA AAAACTGACT
 21661 CATTAACCTC CACGTGGCTT GGAAAAATTA TTTCACTCAT CCAGTAATGA GCTGTTCTA
 21721 GAAATGTTTT GGACATCAAG TCTGTGTTGT TAGCATTATA CATGTTAAGC ATTGAATAAA
 21781 AAACAACATG ATGTGGGTAC ATTTCTTTAC TTACATATAA GTACTTATAT ACTTATAGCT
 21841 GAAAGAGAGG GTTGAAATGT CAGGTGGAAC AGAAATAAGA TTACCTAGAT GTTTCTCCTA
 21901 TGGGTGATTT TCAGCTATGC TGATCTTTCT TCTGGGTCAG GTACTCCCAG AACTTCTTAA
 21961 TTAAATGGTG GCCCTGATCT TAGTTCCTCT CTCCTCTTAG ACATTTTCCA GGACTACAGA
 22021 AGATGTGCAG TTTATAAATG AGTAGCAGAA ACCTACTGAA CAAATTATTC AGGCTCATCT
 22081 GAACAGAGAG GACACCTTCT CTGCTATACT CTCTCAGTGA TTCCCTGCC TTGGGGTCAA
 22141 TTATTGTCTT GGACATTGAT TTAAGCACAT AATAATTGTT GTCATTGCTT ATGTTTGGAT
 22201 TTCATCTCCC AAAATAGATG GTAAATTTCT TAGTTTAGAG ACCAAGTAAT ACTTACAAAA
 22261 AAATTTTGTG TGTGTGTGTG TGTTTTTCT GTGTCTCTCA GCCCTGTAAT AGCATCGTAC
 22321 TTACACTTGT TAGATTTTTA GAGACAACTT TTACAAAACA TGGAAATTAT TACATACCTT
 22381 TTCTACAAAA CAGACAAATT AAATACTCAG TAGTTGAACC AAAAAAGCA GTTCAAATAA
 22441 AATACTTGAA AATGAAGAAA TCATTTGAAC AGAGTTAAAG TTAATCGTAA AATAATGTCT
 22501 GTAAAAATTA TTGCCAATCA AATATAAAGT TCAAAAATAG TGCTTGAAAA AGGAAGAATC
 22561 ATATGAAAAG GGAATACTCA TTTTAAAAAT GTTAGATATC AGGAAAAGCC AAGAAGTGAG
 22621 TATGGTAAGA GTGCTGTCAA GTGAAACCCT GCTAATCTCA CTGAACATGT AAAAATCTGT

Figure 9 (Page 7 of 74)

SUBSTITUTE SHEET (RULE 26)

96/162

22681 AGATGCCTTT ATTTTATTCA CTCACACACA TATGTAGAAA GAGAAATATA TGGTAAACAT
22741 TAAAAAAAC AAATTAGAAT GTAAATTAA TACTTTAAAA AATGGGCTGT ATACTTTTCT
22801 TATCACC GGA GATAAGAATT TATTATTTT AAAATAAAGT TATTTTCTCT GTGACTGTTT
22861 CCATGACTTT GCTACTTAGA AGTTAGAGAT GCCAAAGTTT ATCTAAGAAA ATGTTTATGG
22921 AAATATTATT TCAATAATGA ATGTTTAGAA GACTGAATTT CCTGACTGGG CACAGTGGCT
22981 CATGCCTGTA ATCCCAGCAC TTTGAGAGGC TGAAGAAGGA GGATCGCTTG AGTCCGGGAG
23041 TTCAAGAGCA TCCTGGGCAA CACAGCGAGA CCCTGCAGCA AAGTAAAAAG AAAAAAGAAT
23101 TGAAAAAGGA AGACTGAATT TCCTTTGGGC AAGTCATGTG ACATTCTCTG GCCTCAGTTT
23161 CTTTCATCTAT AAAGTTAATT CCTACATTTT TGGGGAAGGG AGAGAAAAAC TTAGGATAGT
23221 GACTGGCACA GAAGAAGCAC TATATACTAT ATATATGTGG ATATCATTGT TTTTATGGT
23281 ACCATTTTAG CTATCTAATG CAAAATATGA ATCTTTTTTT TCTGGGTCTT AAATTATGGA
23341 ATGTAAGAAT TTTCTAAATT CTCTAATTCT GTGTTAGTTT TAAAGCAATG GAGTAACGTA
23401 TCTGTCAACT TGTAATATA AGGATCAACC TGATCCACAA TTGACCCCT AGCCACTAAT
23461 ATTTAATAGT ACAACACTCA GAAATTATCA AAGGTCAGAG AAGCCAAACA AATGTAAAAA
23521 CATACAGGTG CTCAGAAAGA TGCACCTGTA ATCTCTCTAA GGAGAAATAT TTCCAAACT
23581 GAGTGACACG GTGCTTTAGT GAGTTGTGGA ATCAATCTCA TGATTTCCAA CCTAGTGTTT
23641 TTTTAAAAAT GAACTAGTCC ACAGTAGAAT ATACTAAAGT GCTGGTGCTT AAGATAGTAT
23701 TGTTTTCTGG AAAAAAATA AAAATTTTTT TTTTGTGAGA CAGGGTCTCG CTCTGCCCCA
23761 GGCTGAAGTG CAGTGGCACA ATCATGCTCA CTGCAGCCTT GACCTCCTGG GCCCAAGTGA
23821 TTCTCCACC TCAGCCTTTT GAGTAAGTGG GACCACAGGT ACGTGCCACC ACACCCGGGT
23881 AATTTTTTAA TTGTAGAGAC AGGGTCTTGC TATGTGCTTA GGCTGGCCTT GTGAACCTCT
23941 GGGCTCTAGT GATCCACTAG CCTCAGCCTC CCAAATTTAT GGGATTATAG GCATGAGCCA
24001 CCCTACCTGG CCGTTCCCT GAATTTTTTT TTCTTCAGG TGTGTGTGCA TATGTGTGTG
24061 TGTATGGGTA TAACAGAGAG ACAGAGAGAA AGAACTTTT CTATCACACT TTGCAATCAG
24121 AAGTTTGAAG CTCTATCTTT TGGCTTTTGT TTCAGAAATA TTTCAAATGT AGACTCTCTC
24181 CTTTACCACA CTGTCCCTT AGGCAAGGTC TTTGCCATTC TTCTGAGACT ATTGCAACAG
24241 ACTCCCAACT TCTGACTGTG GGCCCTTCTC AAAAATGATT GTTTATGCAA TAAATCTAAA
24301 CCAAGACAA CTACAACAAT ACAACAAATT CTCTGCTTAA AACTTTCAA TGTCTGCCGG
24361 GCGCGGCGGC TCACGCATGT ATTCCCAGCA CTTTGGAGGC AGAGCGGGG AGATCACTTG
24421 AGGTGGGGAG TTCGAGACTA GCCTGGCCAA CATGATGAAA CCCATCTCT ACTAAAAATA
24481 CAAAAAATTA GCCAGGCATG GTGGTGGGCG CCTATAATCC CAGCTAATTG GGAGGCTGAG
24541 GCAGGAGAAT TGCCTGAACC TGGGAGGTGG AGGTTGCACT GAGCCAAGAT CACACCTATTG
24601 CACTCCAGCC TGGGCAACAA GAGCAAAACT CTGTCTCAA CCAAACCAA ACAAACTTC
24661 TAATATCTAC CAAATGTTTC ACACAAGTAT TTGGGGATCT TCACAAATGG CCCTTATGGA
24721 GTTTTCCTTT GCTGAGACCC TATGCTCTGG CCACACTAAA CTCATTGAGC ATCCAGAAA
24781 GGCCTCAGCC TTTGTGAGCA AGCTCTTATC TCCAGGCCTC TCACAAAGAC CTGTTCCAGT
24841 AGAAGCTCAG GGGAGCACAC TGGACATTAT TCCAACAACC CTTTCCCAC AGCTATGCAG
24901 CCAAATCTGC CAGCTCAGTT AATTAATTAA GCAATTCAGA GATGAGGGTC TGCCAGGCT
24961 GGAGTGCACT AGCTGCGACC TCAAGCTCCT GGGCTCTAAG TGATCCTCTT CAGTCTACCC
25021 AGAAGCTGGG ACTGCAGGCA TGTGCCACCA CACCCAGCTA ATTTTTTTTT TTTTCAGTAG
25081 GGACCAGGCC AACCTAGTCT TGAACCTCTG GCCTCCAGCC TTCCGAAGTG CTGTAATTAC
25141 AGGCATGAAT CACTGCGCCC AGCCAACCCG CCCAGTCTTG TTAGACATGG GGTCTGTAGT
25201 TTCTAGTAGG TTCTTGAGTC TAGGGTTCTT ACCTCATGTT TTATAGTTAA TTTAGGGGAG
25261 GGACTGTGTC TGTTTATCTG GGGATGTAGG GGTGGGCAGG GGGATAGAGG GGAATCTAAT
25321 TAATGAAACC AGAAGCAAAA CTCAGTTGAG GACACCGGTC ATGAGAGTGG CCTGATTATG
25381 GCCAATCTTA CATAATGTGT GAGATCTTGA TATTACCCCA TCCTTGAGAG TCCTCTATAA
25441 AGCTACAGGG ACTTGGGAGC ACCTTTAATT ACAGACAACC CATGTTCTCTG TGGATTATGA
25501 TTTATTAGAT TGCACATGCC TAAATAAAGA CATCCTCTGC AGTCTTTTGA CAATTCTATA
25561 AGCATCTTCT GACTCCGCAA TTAGACAGCT AAGAGATCTG TGTTACTTCC CTCACATATA
25621 TAAATAATTT TAAATAAAAA TCATGGCGTG AATAATTTCT TTCCTCTACC GATTGGAAGC
25681 TATCCATTG GAAGACCACT CTGAAGAGAT GAAATAAGTC TTCTGCCAAA GATTACTTAT
25741 TAATTTACAA GGAAGAGGGG AAGTTTGTCT CCTCTCCGTG AATTTGATTG AAAATCGAGG
25801 GCTTTCTCGA ATAGTTTTGG CATCCAGGT CATTTTTCAT TAAAAAGAGA AAAGTCATGT
25861 CAAATATGAA TTTCCGCAGA TTATTCAGCA CTAGACCCTG GGAGATTCTG TAAAGAGGGG

Figure 9 (Page 8 of 74)

97/162

25921 TTTTGTATA CTCAACTTTT CCGGGTAAAA CAAACACAAA TACTCCTCCT CCAAGGGGCG
 25981 GGGGCGGTGC CTAGGTGATG CACCAATCAC AGCGCGCCCT ACCCTATATA AGGCCCCGAG
 26041 GCGGCCCCGG TGTTTCATGC TTTTCGCTGG TTATTACATC TTGCGTTTCT CTGTTGTTAT
 26101 GTCTGAAACC GTGCCTGCAG CTTCTGCCAG TGCTGGTCTA GCGGCTATGG AGAAACTTCC
 26161 AACCAGAAG CGAGGGAGGA AGCCGGCTGG CTTGATAAGT GCAAGTCGCA AAGTGCCGAA
 26221 CCTCTCTGTG TCCAAGTTGA TCACCGAGGC CTTTTCAGTG TCACAGGAAC GAGTAGGTAT
 26281 GTCTTTGGTT GCGCTCAAGA AGGCATTGGC CGCTGCTGGC TACGACGTAG AGAAGAATAA
 26341 CAGCCGCATC AAAGTGTCCC TCAAGAGCTT AGTGAACAAG GGAATCCTGG TGCAAACCAG
 26401 GGGTACTGGT GCTTCCGGTT CTTTAAAGCT TAGTAAGAAG GTGATTCTTA AATCTACCAG
 26461 AAGCAAGGCT AAAAGTCAAG TTTCTGCCAA GACCAAGAAG CTGGTTTTAT CCAGGGACTC
 26521 CAAGTCACCA AAGACTGCTA AAACCAATAA GAGAGCCAAG AAGCCGAGAG CGACAACCTC
 26581 TAAAAGTGT AGGAGCGGGA GAAAGGCTAA AGGAGCCAAG GGTAAAGCAA AGCAGAAGAG
 26641 CCCAGTGAAG GCAAGGGCTT CGAAGTCAAA ATTGACCCAA CATCATGAAG TTAATGTTAG
 26701 AAAGGCCACA TCTAAGAAGT AAAGAGCTTT CCGGGAGGCC AATTGGAAG GAACCCAAAG
 26761 GCTCTTTTAA GAGCCACCCA CATTATTTTA AGATGGCGTA AACTTGAAA CAAGTTTCTG
 26821 TGACAGTTAT CTATAGGTTT AAGTTGTGAT GCAGCTGAGT TGAAAAGGCT TGAGATTGGA
 26881 GAATTAATTC AGGCCAGGCT TCAAGACCAT CCTGGGCAAC ATAGCCAGAC TACCATCTAT
 26941 ACCAGGGGTC CTCATTCCCC CGGCCACCGA CCGGTAACCG GTCCCTGTCC ATGGCAGGTT
 27001 ATGAATTGAG CCGCACAGCT GAGGGGTGAG CGAACATTAA CCAACTGAGC TCCACCGCCT
 27061 GTCAGGTTAG CTGCAGCATT AGATAGATTG TCATAAGCTC AAAGTGTATT GTGAATGGCA
 27121 CATGCAAGGG ATCTAGGTTT CAGGCTCCTT GTGACAATCT AATGCCTGAT GATCTGAGGT
 27181 TGGAGCAGTT TTAGTCCGGA AATCATTGCT CCCAGCCCCC GCACCCCTGT GTCCGTGTA
 27241 TAATTGTCTT ACACAAAACG GTCTCTGTG TCAAAAAGGT TGGAGACTAC TGGTTTTACA
 27301 AAAAAAGTAA TTAGTCAAGC ATGGTTGGCA CGCTCCCTTA GTCCCTGCAC CCAGGCGTTT
 27361 AAGGATACAG TGAGCTATGA TGGTGCTACC TCACTCCAGC CTGGGTGACA GCGAGTCAGA
 27421 CGTTGTCTCA AAAGTTAAAA AAAAAAAG TTAACAAGA AAAAGGGCTT CTGTGACAG
 27481 ACTGCCGTAT ATCTAGAGGT CCAGGAACCT AAAAGTCTGA TGTCCAATCC TGAAAAGCTC
 27541 GATGGTGCAC TAGAGGAGGC TTTTACATGT AAGAGCATCT AAGTTCTGGA AATGCCAGTG
 27601 TCAGGGAAGG GAAGTGGAGA GCAATTGGC ATCCAACAT AACTTGCTGA TACTTTTTTT
 27661 TTTTTTAACA CAAGTACTAC ATTCTAGTCT TTCTGTGGTG TCATTGTAAC TATTGTTTCT
 27721 TAATATGCTA TCCACTGACT TCAAGGGATC AATAAATAGG AATCAAGGTG TCCCAAGATA
 27781 TGGATTAGGG GAGTTTTTTT TTTGTTGTTG TTGTTGTTGT TTTTCTCTAT TCATTATCCT
 27841 GTAGCTGAAA TTTAGAAATT TCTTCCATTG TGTGTGACTG ATAGAAATAA CAAATTTGTA
 27901 GGTATAGTT GTTGCAAGAA TCTGGAAATC GTGCTTGCTT ATTTCCGAAG TACTATTAGG
 27961 TATATCAACA AAAACACACA TATTACGGTC AAGTGGTTTG ATAATTATT TAATATTATT
 28021 GGTCTAATAC AATTGTAACC CTATGAATTA CTTTAAAGTAT CTTATTATAG AAAAGAATCT
 28081 GTAAGTTTCA TCAAACTACC AGAGCATACC GAAGACTGAA AAATTTAAG AATCCAAACC
 28141 TTAATGGAAA TGTTGGAGGC TGCCCAATTA GGTCTGAAT TCCACCTTCC TGAATCACAA
 28201 ACTTGTTTTA ACTCTCAGTC TGAGGTAAC TACGTTTCTC TTTAAACAGA CATAGTTTAA
 28261 TTTTCTTTG ATTTTGTGATT TAGTATTCTT ACTGATCATC ATAAATAACC AATGCTAATG
 28321 TTAGTCTACT TTGGACCATG GTATTTCGAG AAAGTTTGAA CAAAGTCCCC TGCAAACTA
 28381 TGCATTGCAT TATTTACAT ACATTATGT TTTCCAGACG GTTCAATAGT ACCTCATT
 28441 TCTGAACTTA TTTGTATAGT TTGGCATCTT TTTAAAAATT GTGTCCTATA ATGAAAGGTT
 28501 GTAAACATTA TGTTTTAAAT TTGTATAGAT AAAATCAACC ACAGACCTTT CTTGCTTGG
 28561 ATGTAATTGC CATTGTTC CAATGAGTTC GGAATTAATA GGATTGTGCA AAAATATGCC
 28621 TCACTTGCTT GACATAGCAG AGAGCCATTT TGCCTAAATG CTGTGCCCAG CAATGGACTG
 28681 TCACCAGATT CTCATCACAT ACAGTGAGGA TGAACAACTA GCCTCTCCCA GCAGCTGGCC
 28741 GGTCTCTCAA TAATATGGGA CTCCTCAAG ATGGCTTCCT GCACCTTTGC TCCTCTAGCC
 28801 TTGTATGTAT ACAAGGCTAG CATGCCTGGC ATACATAAGG TTAACAAACA AATCAATAAG
 28861 TTATGGTTCT TCCTCCAGTT CTGGGGATTA TTAGACCACT TTTTGTGTTT GTTTGTTT
 28921 GGATGGAGCC TCGCTCTGTC ACCAGGCTA GAGTGCAGTG GCACAATCTC GGTCACTGTC
 28981 AACCTCTGCC TCCTGGGTTT AAGCAGTTCT CTGGCTCAGC CTCCCACGTA GCTGGGATTA
 29041 CAGGTGCCCG CCACCACGCC CAGCTAATTT TTGTATTTTT AGTAGACGGG GTTTCACCAT
 29101 CTTGGCCAGG CTGGTCTTGA ACGCCAGACC TCGTGATCCA CCCACCTTGG CCTACCAAAC

Figure 9 (Page 9 of 74)

SUBSTITUTE SHEET (RULE 26)

98/162

29161 TGCTGGGAAT ACAGGCGTGA GCCACCGCGC CCGGACTTAG ACCACTTTGT TTTGGCCAAT
29221 AGGACAACAG CCATAGAACC CTCCGCAAAT GAGAGCTTGT CCCTAAAGAT GCTTTATTTA
29281 CATAGCTGTG TGCCGCATGA GCCAAAAGGT GATAACCTTT GTTCAACACG CGCCTCCAGC
29341 CCTTCGGTTA AGTCCAAAGT ACCATTCTTA GAATGCTCTA AAATACATAA TTTTTTTTTT
29401 TTTTTTTTTT TTTTGTAGGA GTCTCTCTCT GTCTCCAGG CTGGAGGGGA GTGGCGCGAT
29461 CTCGGCTCAC TGCAATCTCT GCTTCCGGGC TAGCTGGGCC TACAGGTGCA GACCACCAGC
29521 CCCGGCTAAG TTTTGTATTT TTTTGGTAG AGGGGGTTTC ACCATTTTGG CCAGGCTGGT
29581 CTCGGATTCT TGATCTCAAG TGATACACTA GCTTTGGCCT CCCAAAGTGC TGGGATTACA
29641 GTCGTGAGCC ACTGCGCCCA GCAAAATGCT TTTTGTGGAG CCAATCACTT TATTAGCGCT
29701 TACCTCTCTA TGCTACTTTT ATGCTTTGAA ATTTTGTAC AGTGGGGCCG GTCATGGCAA
29761 ACACAATTCA TTCTTATGCA GGCTGTACAG GTTATTTCTG TCATCCAAAC TCATTCTCGC
29821 AACGCATTTT AGCTCTTTAA ACGACTTTGT GAGCGGCCCT GAAAAGGGCC TTTGGGTTTT
29881 TTTGTTTTTG TTTTTTGAAG TTCTCAGGAG ACCGCGTATT CTAGATTCA GCCGCCGAAG
29941 CCATACAGAG TGCGCCCCCTG ACGTTTCAGG GCATATACTA CATCCATGGC TGTGACAGTT
30001 TTGCGCTTGG CGTGCTCCGT ATAGGTGACG GCGTCTCGAA TAACGTTCTC TAAGAAAACC
30061 TTAAGCACAC CTCGAGTCTC CTCATAGATA AGACCGGAAA TGCGCTTGAC GCCACCGCGC
30121 CGAGCCAAAC GGCGGATAGC CGGTTTTGTA ATGCCCTGGA TGTATCCCG GAGCACCTTA
30181 CGATGGCGCT TAGCACCACC CTTCCCCAAG CCTTTCCGC CTTTGCCGCG ACCAGACATG
30241 ATTCCTATCG CAGTGAAGG TATGAAGTA AACAGTTCCT TAAATACAAA CTTGGCGGAC
30301 CTGATTGAAA ACAACATGAG TTGGCGCGGT TTTTTTTTTT TTTCAAATTT GGTCAACGAG
30361 TGGGTGGAGC AAGAAAACT GTTTCATTAT GGTTCATTGT TTTGATTGGC CAGTGACAGC
30421 TTGCTCTTTG TGGGAGTGA AGGGTGTGTT CAAGTTGAAT GCGCTGTATT CCTGTGAGCT
30481 TAATGACGCT AAGCATAGCC CCATTCCACA TTTCTTTTTA TTTCCACTTG CTAATAATA
30541 AATTACGGAA TAGTTTATTG GGAACATAC AAATAATGTT TAAAGGAGGT CAGATTTATA
30601 GGTCAAGGA TTTACCTCC CAATCATTTT AATATTTTAA TTTAAACCAG GCATTTTGAT
30661 GGCCTTCTCT GTGCTGGACA AGGTATAAGT TTGGCTATGA AGTTTCACTC CTAAAGACCC
30721 TATGTTTTGG GAAGGCAAAA AGGTAGCCAA ATAATTGCAA ATTAAACCT CATAAGTGCA
30781 AACTTCTTCC TCGTCACTTT CCTATCTCG ATTCAAATAT TTGTTGAATG ACTCATTTTT
30841 CTGCAAAAGT CTGAGAGAGA CAGGGAATAT AAACCTAAGT CTGGATAATA TGTTTTCCCG
30901 GGACGCTCTT CCTGGTCTGC TGTGCTGTT TGCTGTGCCT GAAATTCCTA ACACCTCTCC
30961 CTTCCCTCCG TTTTAAATCC CTTTCAACT TGCTACAGCT TTAGAGAAAA GAACATACGT
31021 TTTGTACAGT TGGGGATTAA TTGAAGTGTA GGGCTAATAC TTGATTAAGG TCATTACAAA
31081 ATCTACAGGG TCTTCTCTG GGAGGTTTTT GTGATAAGAT TATTGGTGT AAAATAAGGC
31141 TAATCCCCTT GAAAAATAAA TAGAATAGCA GAATTGGGTC TGAATGTGGT TTGAAGAAAG
31201 GGACTTCTCA ATTCAAATTT TTATTCTTAG CTTCCTGTGG GAGCTTTCCA GAATGCCCAT
31261 AAGATCCACT TTTGTTTAAA AAACAAAAC AACCCACCC ACCACTCTCT GGTAAATAAA
31321 TGAATTTCTA TTGGGAATAT TTAGAATGGG GCTGTGGCCT GTGAGAGACA TTATATAGTA
31381 ACCTCAGACT TGCTCACATG AAGAGAAGAA ATCCAGGAAT GGAGAAAAA GACCCAGGAA
31441 AGGCCAGAAT GCTCTACATG TCATATTGTT TGTATCACTT CTGAAATAAT TGATTACATT
31501 CTTCTGCCCC AAATTGAGTT CTTAGGTTCT TCCACTCACT GTCCACATGC CACAACACAG
31561 ACCTTATAAC TAGAGACTTA GCTAGGAAGA AATGTCAAAC ATTACAGAGA AAAAATGCAG
31621 AGTCTGAGAT CATAAGTAAA ACTCTGAAAT CTCAACATGC CTTTAAATTC ATGAAAATAA
31681 AAAATATAGC AGCATATGCA ATATGATAAT TCTCTGAAAA CATACATCAT GTGAACCTACC
31741 CTGGAACACA TCTCGCCAAG TGCCATCTTC ATTTTAACCA GAGGTCTAGG ATGCCTTTCC
31801 TTTATTTTGC CTATTATATC ATTTATAAAA CCCCATTTT ATTTTGATAT TTTATTTACT
31861 TTCTATTTCC TGCTCCTAAT ATCTCCTTC TAAACTTTTC TCAATGACAG TGAATCAAAA
31921 ACAATGAATG TCAGAACAAA TATTTAAAGG ATCTGTACAT GTAGATATAT ATATTTAAAA
31981 TGGATTCTTC CACTCTGGGA AGAATTCAGG CATACTCAAT CTTATGGTTA GGGAGAGATT
32041 AGGCTCACTC GCCTAATCTG TATGGCTTCT CGTTCGCTTT CCATTTCACT TTCTCTCAC
32101 CCATCAGATC AAACCTATTC ATTGAACAAG AGACCTAAGC CTTTCAGATT AAAACTCTGC
32161 AAACAAGTTG TGGTTGAGAG GATACATGAA GCATTCAAAC AAATAAATCT ATGATATTAA
32221 TCAGAGGTTA ATCTATGATA TTAATCAGAG GTTAATGCAG TGGCTCACGG CTGTAATCCC
32281 AGCACTTCAG GAGGCTGAGT TGGGAGAATC GCTTGAGCTC AGGAGTTCAA GACCATTTG
32341 GGCAACATAG CAAGTCTTCA TCTCTACTTA AAAAAAATA ACCAGAGGTG TTATGAAAAAT

Figure 9 (Page 10 of 74)

SUBSTITUTE SHEET (RULE 26)

99/162

32401 ATAAATTGTC CAGAACTACC CTCCACAAAC TAACTCTCTC AGAATATTCG ATATGAGGAA
32461 TGAATATATGG TGTGTGTGTG TGTGTGTGTG TATGTGTGTG TGTGTGTGTG TGTATGCACC
32521 TATATATGGC ACCTATATAT TCAACAAACA ATTCTGATAA TTGGCCAGGG TTGAGAATGA
32581 CTAGCAGCCC AGCATACACT ATCAGTTTTA AGTATATAAT TGCCTTTAG TAAAATGTAA
32641 AGAAATCCCA GAGTAGAAAT ACTTTTAAGC TATATTACAG GTGAGAAAAT GCATAAGTAT
32701 AGTCTCACC AACTTAGACT ATGGGGGCTT TATAATGTCA CAACAGTTGT TTCCAGGCAT
32761 TTGGGGACAT CACCACTGGT CTTGGGCAAG AAACCTCTCT AGCCAATGGC TGATTTATCT
32821 CACTCCCATC TAAGGCTTCA CTGCATTTCT CTTTTTCAGC AACCTAACTT ATTTAAAAAT
32881 ATCCATTTTC TGATTCATTT TTTTCTGAAT TAAACTGTCA GTACCATTGG CACACCTTTG
32941 GTTCCGTAGC ATACCTGTGT CTCTGCTGTG GTTTTTTTTA CCTCCACTCC TTACTTTTCT
33001 AGAAAAAAT CTCTGCTTTT TCTTTTCAGT TTAAATTATT TCACAAAAAG TTTTCTTGAC
33061 TTGCACTTCC TAGGCTTGCT GTCCTTGTGT GGGCACGCTC CCATAAACAC TATTAATACA
33121 CTTTCGATTG TTAATAATAA AGATATCTGG ACAGAAAATT TCTTTTCTTT TTTTAAGATT
33181 TTAATAATTT TAATGTTTAT TTTTTCCTA GACTGGAGTA CAGTGGCACC ATGATGGCTC
33241 ATGGTAGCCT ACACTTCCCC GGGCTCAAGT GATCCTCCCA CCTCAGCCTC CCAAGTAGCT
33301 GGGACTACAG GTGTGCACAA CCACACCTGA CTAATTTTGT TTATTTGTTT GTTTTGTTTT
33361 TTGAGATGGA GTTTCGCTCT TGTGCCCCAG GCTGGAGTGC AATGGCGGGA TCTCGGCTCA
33421 CCGCAACCTC TACCTCCAG GTTCAAGCAA TTCTCCTGCC TCAGCCTCCC GAGTAGCTGG
33481 GATTACAGGC ATGCATCACC ACGCCAGCT AATTTTGTAT TTTTAGTAGA GACGGGGTTT
33541 CTCCATGTTG AGGCTGGTCT GGAACCTCTG ACCTCAGGTG ATCTGCCCCG CTCGGCCTCC
33601 CAAAGTGCTG GGATTACAGG CGTGAGCCAC CACGCTCGGC CACTAATTTT GTATATTTTG
33661 TAGAGATGGG CTTTCCCTGT GTTGTCCAGG CTGGTCTTGA ATTCTGGGC TTAAGTGATC
33721 TGCCACCTT GTCTCCCAA AATGCTAGGA TTAAGTGGCT GAGCCACCAG GTCTGGCTGG
33781 AAAGATAATT TCTAACATTA TCCTCTCTTA AACATTTGTT TCAAAAATTT TACAAACATG
33841 AGAGTAATTA AATTTGATTT TCAAAATTCC CTTGAATACT TTCTTAATAG CACACAGAAA
33901 GCACAAAGTA TTTTACATTT GTTTTAATG TGAAATTGTG AACCCTAACT TACACAAAGA
33961 AAAACCGTAA CATTATACCC ATACTTAAAA CAGATGCCCT CATATACATA GTAAAACCTC
34021 TGGGGGAGT AGTGAAGTTG GTTATTTACT GTTTTATGAA AGTGCCATT AGCCGGGTGC
34081 AGTGGGTCAT GACTGTAATC CCAGCACTTT GGGAGGTCGA GGCAGGCTGA TCACGAGGTC
34141 AGGAGTTCAA GACCAGCCTG ACCAAAATGA TGAAACCTG TCTCTACTAA AAATACAAAC
34201 ATTAGCTGGG CGTGGTGGTG TGTGCCTGTA GTCCCAGCTA CTCAGGAGGC TGGGGCAGGA
34261 GAATCGCTTG AACCTGGGAG GCGGAGATTG CAGTGAGCCG AGATCGCACC ACCGCACTCC
34321 AGCCTGGGAG ACAGGGCGAG CTCCGTCTCG AAAAAAAGT AAAAAAAGT GCCGTCATAG
34381 TGACTTAGTT TTAAGGAATA AATCAAGGAT ATTTAACTCA ATAGACTACA GTTAGCTAAC
34441 GTGACTTGCA CTGAAAGTTA TACGAATATT GGTACTTATT CCCCTGCCCT TGAAGTATGA
34501 ATTAAGACT CCAAATTTCT TTTTAGAATC TTCAGAGTAA AAGCTAGAA TTGATTTTTT
34561 TAAATAATAA AAAAATACTT TGTATCTAAA TCTGGTGTAT AAAATAACTT GGTGGATGAT
34621 GCTTCAAGGC TATCCATCCC CAAATTTCTC CCTGAATGAT AAAGAGAATA AATGAATATG
34681 TCAATTCAA AGTTAGAAAT TTGGCCGGGC ACGGTGGCTC ACTCCTGATA ATCCTTTCGG
34741 ACGCTGAGGT GGGTGGATCG CATGAGCTCC GGAGTTCAAG ACCAACCTGG GCAACATAGC
34801 CAGAACCCGT TTCAATAAAT AATAGAAAAA AATGAGCCAG GCGTGGTGGT CCCAGCTACT
34861 CAGTAGGCTG AGGTGGGAGG ATCACTTGAG CTCAGGAGGT CGAGACTGCA GTGAGCCGTG
34921 ATCGCAGTAC TGCAACCAG CTTGGTGTG AGACTGAGAC CCTGTCTCAA CAACAACAAA
34981 ACAAGTTAGA AATTTGGCTG GGCGCGGTAG CTCACGCTG TAATCCCAGC ACTTTGGGAG
35041 GCCAAAAAGG GCGGATCATT TGAGGTCAGG AGTTCGAGAC CAGCCTGGCC AACATGGTGA
35101 AACTCCATCT CTAATAAAAA TACAAAAAAA CTTAGCCGTG CATGGTGGCA TGCGCCTGTA
35161 GTCTCAGCCA CTGGGGAGGC TGAGGCAGGA AAATTGCTTG AACCAGGAG GCAGAGGTTG
35221 CAGTGAGCCG AGATCATGCC ACTGCATTCC AGCCTGGGTG ATAGAGTGAG ACTCCATCTC
35281 GAGAAAAAAA AAAAATTCT GTATGAACTG AACAAAATAT CCTTAAATTT TAAAATACAT
35341 CTGAAAGATA TTTCAAAATA TTTAGGAAAA AAATTATAGG GATCAGGCAA ATTCTGAGAT
35401 TCCTTTTTCC CTGCAGCAA CATTAGGAGT GCTGCTGTT CTAACAAACAT GGTAAGTGT
35461 GCCACACCGT ATGTTTCTTT GGCTCAGACA TAAGGTTGTG TAGTTGTTAT TCCAGAATAG
35521 CTAGAATAAA AATCCAGCAC ATCATTTTCT TCAGCAAGTT AACTAACCTC TCTGTGCCTT
35581 GGTTCATAA CAGCAACATA AGCATAACAG AATAGCAGCA ATAGCTCCTA CCTACCTCAT

Figure 9 (Page 11 of 74)

100/162

35641 AAGATTCTTT GGAAGAATTA AATTAAGATT CAGAACACAG CCTAATATCT AGTAAGTAAT
35701 AATAATTGGC TAAAAAAATT TTCTTAAGAT TATATATATT CATGGGGTAC AAGTACAATT
35761 TTGCTACATT AATATATTGC ATTGTGGTGA AATCAGGGCC TTCAATCCAT CCCGGAAAAA
35821 AAAAGTTTTT GAAAAGATTT CTGCCATGGA AAACTTTTAA TGTACAAATT CATCCATCCA
35881 AGAAATAGAA AATATATAAG TATCAACTCC AAATCCACCA TATCTATCTC TTCTGCACCT
35941 TAAACAATTA CTCAGAAATA GAATGCTTGA GATACCAGAA TGCATGCATA TCAAGTAATA
36001 AATGCATGCA GGATGTCAAC GCATCCTAGG CTTTCAAATA AAATTGTCAT ACAAATACT
36061 TTAATATTGT AGTAACATTC TACATGTTAG AGTGTAGAAG TTAATCGCTG ATGCAAAAAA
36121 GGAAAAGAAC ACATTATACC CAAAGCCTAC AGAGAGAATC ACAATTACAA ATATCAGCCT
36181 GCATGTGAAA ATCTTTAATT TGAAAGTCAG AAATATTTAA ATGATAGTCA TTGTTAAATC
36241 AGATTGTGGT TTGAAAAAAA GTTAGTTTAA AACTGAGTTT ATGAAAAATT TGGGGATTTT
36301 AGAGACAGTG TTTTGTTTTT AAATGTGTGT GAGTTTGTGA AGAATGTTTT ATAAAATACT
36361 GACAGTATTA TAAGATGACA TTATTATAAT ACAACATAAG AATTTTGGCC TGTACCTCTC
36421 AGCAGTCCTC AATCACCTGC TGTACTTGAC TCAATGATTA TCAGAGTGGT TTGTTTTCTC
36481 TCTGTTGTGT TCCCAGTTCA GGCAGCTCAG CAATGGCCTG TGATTCCAGC AATTCAAATA
36541 GCTGGTAAGT AGTTTCTTGT TTGTTTTCTC AAATTTTCAG GGGCTTTTCT CTACAAGTGA
36601 TTTCCAGTGC ACGCCCCCTC ACCCATCTCT TATTCCTTTA CCTTCAGGAA AACCCCTCAGC
36661 GCTGCATCTC TGGTCACCGG ACCACCGTGG TACATTTACC TATGGCCACC AGGTGTCACC
36721 CTTCTCTTTA CTACCATGGT TTGTGAATGG TTTTGCCAGA GGTGAATAAG AATTTAAAT
36781 GCAGGTCTTT GATTTTCAA ATGTAGTTGA CCTTAAGAAT TTATGAATAA AGCCAGAAAA
36841 ATTAAGCTTA AAAACACCG AAAGAAAAATG AGGACTTAA ATTTCTATTA AAAAAATTAA
36901 CAGGCCACAG TTGCTGATGT TTAGTAAATG TGTTAGTGAA ATGTGTTACT GTGAAGACTG
36961 GGGTGTCTCT TGAAATCTCA GCCCAGGTGA AATAAAACCA ATATAAAACA AATGCTTACC
37021 TAATAAATTA ATTGTAACAT ATTCCTTATG AGGTAGAAGA GTAAGTGAAG CCTTATAGCA
37081 GTCTGCTTTC AGTATAGTAA GATATTAAGA GAGAAATAAT TTGTCATATG CTTTCAGAAT
37141 GGTTTGCTGG TAAAATAACC AATGTCTTAC AACTTAGACG ACAATGTCCC TAGAGTGAAG
37201 AAACACGATT AATTCGGCTA CCACAGTTGA ATGAAAATAT TCCGTAAGAC AAAATGTAAA
37261 GAAATTAGAA GCAAAATAAA TGTCTCCAAA ATGACAAAGC GATTAAGTAT ATACACAAGA
37321 TGAACAAGAA CTCAATAAAA ATCATGCAGT ATACAATACA ATGTACATTT ATTAAGTAT
37381 ATGCATTTTT AATGCAACAA TAATACTAAC AGGTAATAGA CAAGTTGTTA ATAGTTTTTC
37441 ACTGGCTAAT TAAATAACAG CTTTAATTGT ATTCATTTTA TAGCTTTTCT ACAATGAGCG
37501 TAAATCACAT TTACTTTTTT CTACATAACT TTTCTAACCA CAAAAAAGA AAATGGTTTA
37561 AAAGAAGAGA TGAGATATCT TTGCTAAAT TTAATGCCTA AAGAAGAAAC TTCTGAGCTG
37621 TATATGGTAT CCTGAAGCAC CTGCCCTTCA AGACAGAATG CTTGTACCAC ATTTATGCAG
37681 CCAAGTGCAT GTAGTAACAT AAAGTAAACA CATGCCATCT GGATATATAT ATTAAGACTC
37741 TTTTGACGGC TGGGCAGGGT GGCTCACACC TGTAATCTCA GCACTTTGGG AGGCCGAGGC
37801 AGGCGGATCA CGAGGTCAGG AGAGTTCGAG ACCAGCCTGG CCAACATGGT GAAACCCTGT
37861 CTCTACTAAA AATACAAAAA TTAGCCGGGC ATGGTGGTGC ACGCCTGTAA TCCCAGCTAC
37921 TTGGGAGGCT GAGACAGGAG AATCGCTTGA ACCTGGGAGG CAGAGGTTAC AGTGAGCCGA
37981 GATCATGCCA TTGCACTCCA GCCTGGGCAA TAGAGTCTCA AAAAAAATAA AAAGACTCTT
38041 TTGAACATGG TGAAGTGATT TCCCAGAATC TAGCAATTCC TGAATGTCCT GGTTAGATT
38101 TTTTTTTAAT GTGCACCGGA ACCCCAGTGG CTCCATGGAA GGACCTGGGC ATCCTCTAAG
38161 CCACTTGGTG GCTTCCATTA TACCATCTCA AAATGAGAGA GCTTACTCCA CTTCAATTGAG
38221 GGAAATACCA CCAGAGTTCT GACTCCAGAG GCACTGGCCT AGGGAGGACA CCGTGTGTGA
38281 AGCCCAGCAG GGCCACTAGC TGTCCCCACC AATTACAGTC CTTGCGTAGG GTCCAAAGAA
38341 ATGAATGCCA AAGAGAGCAA CAGAGGAGCA AGGGAGTCAC ATCCAGGAC CTTCCCTCAG
38401 GGACTTTTAA AGGAAACATG ACAGCTGAGG ATCAGTTGGT TGTTTTCTGC TGTTCCCTT
38461 CATGTGATTC AAGCTCATTC AGAAGAAACA CAATGAGACA AGAGAAGAGC CATCTCCTTC
38521 CTTCTCTATT TATTCTAGGC ATCTAAACTA CTGAATGTAG TGGTGTCTGA GATGTATCAA
38581 ACGGTCAGAT TGACTGAGTT TGAAACCTGT TTCTATCACT GACAAACTAT GAGATACTCT
38641 ATACTTCACT TTCTTTTTT TTTCATTTTT TTATTTTTAT TTTTATTTTT TTGAGATGGA
38701 GTCTCACTCT GTCACCTAGG CTGGAGTGCA GTGGCGCAA CTGGCTCAC TGCAAGCTCT
38761 GCCTCCTGGG TTCATGCCAT TCTCCTGCCT CAGCCTTCCG AGTAGCTGGG ACTACAGGCG
38821 TCTGCCACCA CGCCAGCTA ATTTTTTGTA TTTTATTAG AGATGGGGTT TCACCATGTT

Figure 9 (Page 12 of 74)

SUBSTITUTE SHEET (RULE 26)

101/162

38881 AGCCAGGATG GTCTCGATCT CCTGACCTCG TGATCCACCC GCTTTGGCCT CCCAAAGTGC
38941 TGGGATTACA GGCGTGAGCC ACCGTGCCCG GCCTACTTCA CTTTCTTCAT TTA AAAAAGA
39001 AATGGGGATA ATAGTACCTA TCTCATAGAA TTATTGTAAG AAGTGCATGC AGTAATGCAT
39061 GTAAGTAGGT GCTCAGAAGA GTCGGACACG AAGTAAAGTGC TTTTATCATC CTTATCATAA
39121 TTTTCATTAT CAGAACAAGG AGAGACCAGG TAGAAAATTA TTGTGATTCT TCAGGTCCTGG
39181 AATACTAGAG TAGCATCCCA AATGAAGGCA CCATTAAACT TTGCAAATCT GTATGACACC
39241 TTCATGCCAA TTAGAAAAAA CACCTCTTCA CAACCCCTTT CAAGATATTT GCCTCCTACC
39301 TGCTAAAAAC ACCCATCATA CTACCCACAG ATAGCCATGA TGCTTTTTCT GGGACAGGTG
39361 CCTCTTCCAT TCGTGCAGTG TACAGCCTTC ATAGCTGTGC AACTCACATC ACAATCAGAT
39421 GGAAGAATCC CCAAGGCTTG GTGACAGATG AGTTACTGGG TAACACAGAG AGAGGATTCA
39481 AAGGAAAAGT TGAACGGGTC CAGAAAATGC ATAGATACAT GTGTAAAAAT CTGTTAAGGT
39541 TATGACTAGC CACGTCCCAG GGTTCAAAGC TTTTCTCAGA TGTAAAAATG AATCATGTAA
39601 GTCCCCCAA TTTAAGGAGT CCTCTTCCAA AAATAGGAAA TGAAATGACA TAGGTGTATG
39661 TCTCTGAGGT GACGGAGGAA ATGAAGGAAG CCTCTAGATG CAGCTTGAGG TTCATGAGAG
39721 ACAGTTCCAG GGGAGAGGTC ACAGCTAGGG ATCACC GGCA TGCAGGA ACT CAGAAACCTA
39781 AATGGGGAAA TCTTTTTGAG GAAATGAACA GAGAAGGCTA AAATCAAGGA GTTCGTCAGG
39841 CAATTTCTAT GTTTAGGTTT AACTCTCTCC TGAAACATGA AGAGCTCATA AATGCACTCC
39901 CTCTTTGAGT CTCTAGTTTT GTCTCCTTCC CACAGTGAGT CTGCAGGCTG CGTGTCACTC
39961 ACGTTACAGT AAGACGTAGT GCCCCATGGC TCCTCCTGTG GAGACAAGAG ACCCAGGAAA
40021 GAGGCATCAC AAACCTAGGC ACCATCTTGC CTCTTCTCTC TTCCTTATTT TCCTCATTCA
40081 CCCATCTCAA TTTAGACCTG GGCCTATTG GATTTCAAGA ACCATTATCT CTCATCTGGA
40141 AATGCTTATT GGCTTTCTAA CTGGTCTCCT CACCTCTCAT CTAACCTCTT AACAAACAT
40201 TCACCATATA AGGGAGATCG TGGTCTCCTT TTCTTAGGAT CCTTCAATGA CACCCAGTG
40261 ATCATAACCC AATATCCCAA AAGACCCTTG GACTCTGTAT GAGCTGGCTT CTTTCTGATT
40321 CTCTTTTCCC TACACCACAG ATGTTACAGG GGTAGAAATG CATAATTGGT GAGTGATAGC
40381 TAAGCAA ACT CAGGGTTAAG GTACAGTAAT TATTTCTAAT CTCCCAGTAT GCCTTATACT
40441 CTCTACTTG GCATGGTTGC TCCGTCTGTG TAGACCTCCC ATCATCTTCA ACCTCACCTA
40501 ATGGAATCCA GCTTCTCCTT CAAGATCCAG AAGGCTATCT TGATCCCCAG CTGAATGTGA
40561 TCATTCTTTC CTTTGACACC CTAAGCATTT GCTTCTGCC TGCTTTAGGA CCTCATGGGG
40621 TCTTCTTTAA CTACATTAC TTGCTATCAA TTTTCTTCCC TACCAGATTT GGGTCTCTGAG
40681 AATAGCCACA GTGACTTCTC AACCTCAAAG CCCTGTACT ACCTTAAACA GCTCTTGCAA
40741 AATAGTAGGT GCTCTGAAGA TGTTTGTTGA ATTAGAGACT TTCATTCTGG GGAGAACCAT
40801 TATTTTCTGT CTCCAGGGA GCTGCTGGTG TCCCCAAAGA ATATAAATGA GAAAAATGCT
40861 TCCCATGGAT GCCAGATCCC CTCTGCCCT CTTCCCACTG TGCCCTGGGG CAGAGGTACT
40921 AAGAGACTTC CCCCTTGTTT CTA CTCACTT GAACCCCTGCC TCTTCCTTAA TATTATGAAC
40981 AAAATTCCAA TGAACAAGAT GACGACAAA ACAGCAATTC CACTGATGAC TCCAATGACT
41041 AGGGTGCCAG ACGGTGAGGG CTCTAAAACA GAAAAAGCAA GTTAAAGCCT TTGATTGCCA
41101 CCCTCAGCCC ACCCCCTAAC AAAGAGCAGA TCCTCATCTC ACTGCCATAA TTACCTCCTC
41161 AGGCCTCCT CTCAACCCCC AATAGATTTT CTCAGCTCCT GGCTCTCATC AGTCACATAC
41221 CCCAGATCAC AATGAGGGGC TGATCCAGGC CTGGGTGCTC CACCTGGTAC GTATATCTCT
41281 GCTCTTCCCC AGGGGGTACA GCCAAGGTTA TCCAGCCCTG GTAGGTCCCA TCCCCATTGG
41341 GCAATACGTC TTTAGGTTTC AACTCCTTGG CATCCATTGG CTGCTTATCC TTCAGCCACT
41401 TCATGGTGAT GTTCTGGGGG TAGTAGTTCA AGGCCCGACA CCGTAGAGTG GTCAGTGAAG
41461 AGGTCACATG ATGTGTCACC TTCACCAAAG GAGGCACTTG ACAGGAAAGA GGAAGGATGA
41521 GGAGAGGGGA TCTGTTTACC CTTGCCAGGA AGACTGGAAC TTTCATTTC TTCTATAGGT
41581 TGGAGGAAGG AAATACCCTT TTCAGAAAAA AACAAGCTAC AGGAGAGACA CCATTTTGTG
41641 TCCTAAGATT GGA CTCTAAC ACAGTGTAC TTGGAGAGCA GTCAGATCAG CTTGTTCTCC
41701 TCACATGTAA ATATACATAT CTGTTACCCA TGTTCTTTGT TCTGATAGAT AAAATTGCC
41761 TTTATGTGCA TTGAAAATGA TTGAATACAG ATGGTCAGTT TCACCTGGGT CAACCTAGGA
41821 GGCATTGTTA TAAGAAGCGG ACTTGTAAGA TAGGTAGCTT CAGTGATTAT TGCTATGTTT
41881 TATGAAAGAA ACTTTTAACT TAAAGGATTC TTCTACTCTG ATAAGTGGCC TCAGTTGATA
41941 TTTTGTCTG GTATTCATAT GATAGCTGAG ATCTCTGAAT TCTCTTTTTT TTTTTTTTTT
42001 TTTTAAAGAT GGAGTCTCAC TCTGCTGCCT AGGCTGGAGT GCAGTGGCGC GATCTTGGCT
42061 CAGTGCAACT TCCGCTTCCC AGGTCAAGC GATGCTCCTG CCTCAGCCTT CCAATTAGCT

Figure 9 (Page 13 of 74)

SUBSTITUTE SHEET (RULE 26)

102/162

42121 GGGACTACAG GTGCGCATGA CTGTGACCAG CTAATTTTTG TATTTTTTTA GAGACGGGTT
42181 TCACCATGTT GGTCAGGCTG GTCTCAAACCT CCTGACCTTG TGACCACCCG CCTCGGCCTC
42241 CCAAAGTGCT GGGATTACAG GGGTGAGCCA CCGTGCCCGG CCTTGACATT TCTGAATTTT
42301 TAACAGGTAT AAATATACAA AAGATTATTG GTTAAATAAA AAGCAAGGGC CATAGACACT
42361 TCCCTTTGAG CCATATGCAT GGAGAAAAGA AATTAAACCC ATGACTTGTG GCTGTCTCAT
42421 ACATCTCAAT TATAAGGTAG AGACTCTAGG ATTGAGAAAG TCCCTTCCCA GAATTTGGAG
42481 AGGCACACAG CCTCAGCCAC CTCTGAAACT CCAACCAGGG ATTCCGTGCC CTGCAACCTC
42541 CTCCACTCTG CCACTAGAGT ATAGGGGCGAG AAGTGTGTTT CCACCATACC TTGTTGGTCC
42601 AAAACACCTC TCCCCAGCTC CAGCAACTGC TGCAGCTGTG CAGGGCAGTC CCTCTCCAGG
42661 TAGGCCCTGT TCTGCTGCTC CCGAATCTTG TGCCTTTCCC ACTCCAGCTT GGTGGGCCAG
42721 GCCCTGGGTT CTGCTGCTCT CCAATCCAGT GTGTCAGGGC AGAATTCAGG GTGGTCTCTG
42781 CCATCATACC CGTACTTCCA GTAGCCCTCG GTACTGTTGT CTTCTTGCAAT TTCACAGCCC
42841 AGGATGACCT GCAGGGTGTG GGACTCTGGA AAAATCCCCA GCCTTGTTAA CTGCAACCAA
42901 AGGAATAGGT CCCTATTTCC ACCATCCCCA AGGACCAAAT GATCTCAGGA AGCAAATTC
42961 TTCCCTCTTC CCTGCTCCCA CAAGACCTCA GACTTCCAGC TGTTTCCTTC AAGATGCATG
43021 AAAAGATGAA AAGCTCTGAC AACCTCAGGA AGGTGAGGCC CCCTCTCCAC ATACCCTTGC
43081 TGTGGTTGTG ATTTTCCATA ATAGTCCAGA AGTCAACAGT GAACATGTGA TCCCACCCTT
43141 TCAGACTCTG ACTCAGCTGC AGCCACATCT GGCTTGAAT TCTACTGGAA ACCCATGGAG
43201 TTCGGGGCTC CACACGGCGA CTCTCATGAT CATAGAACAC GAACAGCTGG TCATCCACGT
43261 AGCCCCAAGC TTCAAACAAG GAAAGACCAA GGTCTGCTC TGAGGCACCC ATGAAGAGGT
43321 AGTGCAGAGA GTGTGAACCT GGAGACAGAG CAACAGGCCT TAACCATGTG TAGTAGGAGG
43381 GGAGCAGGAT GTTGAGGCTC CACACACCTG CATCAACTCA TACCATGAGC TGTGTCTGGT
43441 CCTCATTTTG TGAAGGGTGA GTTGCACTCC TGTCTTTCTT CCATATGACA GTCCTGGGTG
43501 CTCTTTCCTT GTGTGCTTTT CTCTGCCACA CGTGGCTGCC ACCCCCTCAC TGCCCCCAGA
43561 TCCTATTCCA ATACTCATGA TTAGACAGAC TCCACTAAAG CTGGTGGATT CTAGAAAATG
43621 TTAAGGTGTG TCTAGCCATG GTAGTTGAAC TCAGGAGTTG GTGCTCAGGG CAAATTAGAC
43681 CCAAATCCTG AGGAATAATT CCTTCAGTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT
43741 GAGACAGAGT CTCACCTAT CACCCAGGCT GGAGTGCAGT GGCACAATCT CAGCTCACTG
43801 CAACCTGCAC CTCCTGGGTT CAAGGATTTC TCCTACCTAA GCCTCCTGAA AACCTGGGAC
43861 TATAGGCGTG CGCCACCACA CCAGGCTAAT TTTGTATT TTAGTAGACA TGGGGTTTCA
43921 CCATGTTGGC CAAGCTTGTC TCAAACCTCT GACCTCAAAT GATCTACCTG CCTCAGCCAC
43981 CAAAGTGCTG GGATTACAGA AGTGAGCCAC CGTGCCCGAGC CTGCTCCTG AATTCTTACA
44041 CTGAACTGCC TATGTGGCCT CACCACTTGG AAGCCTGACT GGAATCTCAA ACTTAACATG
44101 TCCAAATGCA GATCCTTGAT TTACCCCAAA CTGCTCTTTC CTCTGCCTTC ACCATCTCAG
44161 AAATGGCATT GCCAATTACC CCACTGCTCA GGCCAATAAA ATTAATAA AGAACAAAGT
44221 CAACTTTAA TCTTCTCTT TTCAGGGGGT CAGGGGAGAC AGGGTCTTGC TCTGTCACTT
44281 AGGCTGAAGT ACAGTGGCAC AGTCATGGCT CACTGCAGCC TCAACTTCTT GGGCTCAAGC
44341 AATACCCTCC ACCTCAGCCT CCCGAGTAGC TAGGATCACA GGTGCATGCC ACCACACCCA
44401 GCTAATTTTT GTATTTTTTG TAGAGAAGGG GTTTTGCTGT GTTGCCAGG CTGGTCTTGA
44461 ACTCCTGAGC TCAGGAATCT GCTCTCCTTG GCCTCCTCCT TGGCATGAGC TACTACACCC
44521 AGCCAATTCT TCTCTTCTC TCACACAACA TAGAATCCTT CAGCAACTTC CTTCAGAATA
44581 TATTCAGGAG ACAATGGTTT GTCACCTCCT TTTCTGTTCC CACCCAGCCC ACTCCACTAC
44641 CTCTTGCCCTG GACTGTGTAA CAGCTTCTTG GCTGGGCTCC CTGCTTTTAC TGTGCTCCCC
44701 TTCATTCTGC TTTCCACATA GCAGCCAGAG CAATCTTTA AAAGCCTGTG ACAGATCACT
44761 GTTACTCCTT GGCTAGAATT CACACCACAG CCTACAGGCG CCTGCACAAC CTTGTTTGTG
44821 GCTCCTCTTC TGAGCCCAAT ACCTACTTCT TGGCCTCTAC TCCCCAGCAC TACTTGTTTA
44881 TTTTTTTCAA CCCGAGCTTC TTAACCAGGA GTTTGTCTAC TAGGTGACAT GTGGCAAAGT
44941 TTAGAGACAT TTTTGGTTGT CAAGACTGGG GGAGTGCTCC TAGCACCTAG TGAGTAGGGA
45001 GGACAGGATA CTGCTAGACA TCCTACATGC AGATGGTAGT CCCCCTTCCC ACCCCACGCG
45061 CGCCCCCCCC CCCACACACA CACACATGAG TAGTGCTGAG AAAACCCGCT TTTTAATCCA
45121 ACTTGCCAGG CCCACTCAGT TTGCCTGGGA AATACTGCTC CCAGTCAATA TCATTCTTAT
45181 TTCCTTCATG TCTCTGCTCA AGTGTCAGCC CCAGAGTGAC TTGCCCTGAC TTCTCTGCTT
45241 CTCACAACAC CCATGATTTT CTGATGTTGT ATATCTTTCT GCTCATTTGC TTATTGTCTAT
45301 CTCTCCCACT AGAATGCAAA ATATCAAAGG GTAAAGACTT GTTCCCTGCT TCTCTCCCTT

Figure 9 (Page 14 of 74)

SUBSTITUTE SHEET (RULE 26)

103/162

45361 GGGGCTTGAA CAGTGCAACA CATGGCTGGG ACTCATTAC ACTTGTAAC AATGAATATT
45421 TCTGCTCAAC ATGAAATTTT ATTATTC AAC CTCTAATGCA GTGTGATGTT TAAGAATCAT
45481 AGCTATGAAG TGGAGACATG AGCTCTGCCA CCAAAGCCCC GTGTACCATT GAATAAATTT
45541 GCCAGGAAGC AGGCCGTGCC ATGCTTCATT CTTGTCATGT GTAAAATGTG GATACACGTA
45601 GTACCAAAAC TCAAAGTGCT GTGCTGAGGC CGGCGTGTGA CCCACAGAAC ACTGTGCTAC
45661 ACTACAGGGC AAAATCACTG TCAACTAAGA TTAGAAGCAG CTGTAGTACT TGAATAACA
45721 TCAGAAAACC AGATTATTTA TGTTCTTTGT AACCTGAAA GAGTTATATA ATCTGAATTC
45781 CAGTTAACTT CTAGTAAAAT AAACGTATTA TTAGCTCCTA CCTCCCTATG CCTAGTGAAA
45841 ATCAAATAAG ATCAGATATG AATGTAACCT AGAAGTGAGT GCATTGCTTA CATGTTTCATT
45901 ATCAGTACTT TGTAGAGAGG CCTCTTAATT ACACAGCACA TTGCAAATCA ATAAAGCCTA
45961 GCCGAAAAGA GAATTGTTCA GTTCAAACGT TCAAAACTAA CATATACTTA ATTTTCCAGG
46021 CAAAAGAACA ATTGCCAAGA GTGGGGAAG GCGCGTGTGA CAAAAGTGGG TGAATGGTG AAGAATTACAG
46081 CACCCTAGAG ACCTCCACCC CAGGTCTCAC CAAAAGTGGG TGAATGGTG AAGAATTACAG
46141 ATCCCCAACG CCACTCTTTC GCGCCCCCAC CGCCCAACGC ATTCGTTCTG AGGTGGAAC
46201 CCCGTGCGGA TCCTGCTGTG GGTGCTGCTA GCCTTCTCGG CAAGCACTCA GGAAGAAGT
46261 TCCTGTTTGG AGATGACTGG GGAAGAACT GCACAGCTGA CATTGGAAT AAACCCGAGT
46321 TCCAGGTTCA AGGAGCCCCA GGCTTAGCTC AGCTCAAGTG AGGAACTACG AGATTTATTT
46381 AAAAGCATT TAGTTGGGGG AAGGGAGTGG GCGGTTCCAA AAGTCACTCC GCAGAGCCGG
46441 GACAGCCGGG GGAGGGGGCA GGTCTGGGG CGAGGGACCC CTATCTGCAG TTCAGTGGTA
46501 GGCCTCCCT CACGGGTCTT GGAGCGAGAA AGTAGGGAGA GGGGCTTGGC GATTGGGGTTG
46561 AGCAGGTCCT CCAAAGTTAG CAACTCCCA AGCGCAAAGA AAAAGCTAGT TTCGATTTTT
46621 CCACCCCGC CGCGCCCTA GTTCGCCGC AGCCCTCGGA CTCACGCAGC AAGCGCCCT
46681 GCAGGACCGC GGTCTGCAAA AGCATCAGGA GGAGAAGCGC CGGCTGGCT CGCGGGCCCA
46741 TTTCCCCAGC TCTGGCCGCA CGTCCCCGTT AAATCTCCGC TTCTTTTGGG GGGCGGGGAA
46801 ACGGGGATGG CTCCAGAAAT CACCCTACAG CTATTGCCTA GGCTCAGGAG ATGCCAGTA
46861 AAACCTCCTG GTGAAAAGCA ACAGGTCTTT CAGAACTTTA GTTCTCTCTC TCCTACAGCA
46921 GAAGGTACCT GCTTGTGAAA CACTAGGTGA TCCAGTGTCC CCCTTGGTTT TTAAATCCTG
46981 AAGGGGTGTT GTTGATTGGG GAAAGTAGCT TCGCAATGTT CTGATCTGAA CTTTAGATAT
47041 TTAAATATTT ATGATTTTCA AAATCAATC ATACATTTAA AAATTTTATC TCAACCTTAG
47101 ACCAACTTAT GTCTTATTTG ACTTAGAAAT ATAAAGCTTT TTCATTTTGT TTTTGTATTC
47161 AAATTAATTA AGTCATAACA TTAACCAATT AGATCCTACT GAAACACGTT CCACAGCCTT
47221 CATAATTGAA TTATCTGACA AGTGTTCAC AAACCTTACA GTATTGGGAT TATCTGGAGA
47281 ATGATTAAAC ATATTGAGGC CTGCTCCTAA CCCCAGACAC ACTGATTTAA TGGGTAATTG
47341 TTAGGTAGTT AGACATTAGC AGTTGGGAGG GGATGACAGA AGAGAGCGGA AAGGCTGTCA
47401 CTAAGACAGC CACTGGCCCA CCTAAATTCA GGCCCAAGAC TACCCTAATG CCACCTAAG
47461 GGATGGAGTT TATGATAAAG TCTGTGGCCA AAATATCCTG GAGAAAGAGA AAGGAGGGTA
47521 CAGGTGGAAT TTCCCTAAGG TGGCAGATGC CCAACAACAC AAAAGCCTGT CTTCAAGTTC
47581 ACCCCAAGTT CATCATGCCA TCATTATAAT AGAATTTACA TACAGTTTGT CCCCCCATC
47641 CCTGGGAGGC TTTTCTTAAC AAATTATAGG TAAGACCATG CACAGTTTAA TTTTAGATTG
47701 TATAGCTATA AACTTCAATC AAATAACATC ATCCTGTCAC TCAGATACAG CCCAAACCTC
47761 AACTCCTCCC CACAAACCCC ATAAAAGCAC CTTGAGCTCT GTAAAGAAGT GCTGAGTTCA
47821 CTTGCGAGAA ATAAGCCCGC TGTCCTCAG AGTGTATTAT TGTGCTTCAA TAACTTTGC
47881 TTTAAGCTTG CATTTTGGTG TTAGTTTGTA GTTCTTTGCT CACTATCACA AGAACTGAGA
47941 TTGCTGCTTC AGAGCTCCGG CTATAATAAT CTCCTCGGT AAAGGATCCA TCCCAATGCA
48001 TAATTCAG TAACAGTATG GGATGCCACC TGGGCAATGG GATTTTAAAA GCTTTCCTTC
48061 TCCCTCAACG AAGTTTGGGA ATTATTGCCT TAGACATTC AAACAATATT AATAAATTTA
48121 ATACACCTGA TTTGCTCCAA ACCTTTACAT ATCTAGCAA TTCAACAGGC ATTATTTTGT
48181 TAAGCATGTA TGCAAATTTT GGCAATTCAA GAAAATCAA CAGGATATCA GGGCCTCGAC
48241 TGTAGGCAAA CAGATACAAT AACATTGGAA ACATGTAGAA TATTGATGAT GGGCACCATTG
48301 GGGCTGATAG TACTATTCCT TTTTTCAT TTTTGGTAAG ATATAATTAG CATACCATAT
48361 AATTCACTTA TGTAATATGC AAAAATTGGC CCAGCTCAGT GGCTCACGCT TGTAATCCCA
48421 GCACTTTGGG CGGCCGAGGA AGGCAGATCA CCTGAGATCA GGGGTTCCAG ACCAGCCTGG
48481 CCAACATGGT GAAACCCCGT CTTTACTAAA AATACAAAAA TTAGCCGGGC GTGATAGCAG
48541 GCAACTGTAA TCCCAGCTAC ATTAGAGGCT GAGGCAGGAG AATCGCTTGA ACCCGGGAGG

Figure 9 (Page 15 of 74)

SUBSTITUTE SHEET (RULE 26)

104/162

48601 CGGAGGTTGC AGTGAGCTAA GATCGTGCCA TCGCACTCCA GCATGGGAGA CAAGAGCAAG
48661 ACTTCATCTC AAAAAAAAAA AATTAGCTGG GTGTGGTGGC ATGCACCTGT AATTCCAGCT
48721 ACTCGGGAAG CTGAGACAGG AGAATCGCTT GAACCTGGGA GGCGGAGGTT GTGGTGAGCC
48781 GAGATCATGC CATTGCACTC CAGCCTGGGC AACAAGAGCG AAACCTCCGTC TCAAAAATAA
48841 AATAAATAAA ATAAAATGCA AAAATTAATG GATTTTAGTA TATTACAGA GATGTGCAAC
48901 CATTACCAA ATTTTACATT TCTATCTCCC CAAAAAGAAA CCATGTTCCC CTAATTCAGT
48961 ACCCTTAATT CATCGCCTCC CAGATTCCTC CATTCTCCTC CTCCTCCCCC CCCAGCCCTA
49021 GACAATCTTT AATCTACTTT CTTTCTATTT GGAACATTTA GTATACATAG AGGCATATAA
49081 TATATTGCTT TGCCGTGACT GGCTTCTTTC ATTTAGCATA ATGTTTTTAT GTATGTTTTT
49141 CATGGACCAA TAATATCTAT TATAAGGACA TACCACAACA TATTTTATTT ATTCATTCAT
49201 CAGCCGATGG ACATTGGTTT GTTCTACTT TATGGCTATT GGAATAGTG CTGTTATAAA
49261 CATTTATGTA CAAGTTTTTT TGTAGACTTA TGTTTTGATT TCTTTTGGTT ATATATCTAG
49321 AAGTGGGTTT GCTGGGTCAT ATGGTAACAC TGTTTAACCT TTTGAGGAAT TGCCACATTC
49381 TTTTCCAAAG TAAGCATTTT ATCCTCCTAT CAGCAGTGTA TGAGAGTTCT GATTCTCTC
49441 CATCTTTGCC TGGGTTTTTG AATCAGGGCC CCAGATAGAA CAAAAATGTG GTTATTCAGT
49501 TGTTCCACCA TCACTTGTTG AGAAGACTCT TTTTTCATTG AAGTGTGTTG GCACCCCTAT
49561 CAAAAATCAA TCTACCATAA ATGTGAGAGT TTATTTCTGG AGTCTCAATT TTATCCCAT
49621 ATGCTATAAT CTATAATCCT ATCTTTTTTT TTTTTTGACA GAGCCTCACT CTATTGCCCA
49681 GGTTGGAGTG CAGTGGCCCA ATCCCGGCCA CTGGCTCCTC CTCCAGGTT CAAGCAATTC
49741 TCCTGCCTCA GCCTCCCAAG CAGCTGGGAT TACAGGTACC TGCCACCATG CCTGGTTAAT
49801 TTTTGTATTT TTAGTAGAGA CGGGGTTTCA CCATGTTGGT CAGGCTGGTC TGGAACTCCT
49861 GACCTCAGGT GATCTGCCCA CCTCAGCCTC CCAAAGTGCT GGGATTACAG GCATGAGCCA
49921 CCACACCCAG ACTATAATCC TATCTTTATG TCAGGACTAC ACTGTCTTGA TTACTATAGC
49981 TTTTGTAGTA ATTGAATTCA AGAAGTTTCT CAACTTCAAA TTTGATCTTT TTTTGAAGA
50041 CTATATTAGC TATTCTCAGT CTGCTGAATT TCCCTAGGAA TTTTAGGATC TATTATCAAT
50101 GTCTATTCTA TTTTGTATA TGTTTTAATA TTTTCATAAG AAACTTTTTT CATTTAACT
50161 TTTTTTTTTA AGAAAAATAG TGAAATCAG AATACTGGGG GTCAGGCGCA TTTAACAGGC
50221 AGAAGAAGAA TAAAACTTG TCATATAAAC AAAAAAGAAA TGACCAATCA CATTGTGGAA
50281 GCCATGGAGT GGTATAGGT GCCAAGGCT GCAGAGAAAT GGTGTCAGAT ATACCTGAAA
50341 ATTGTCCATT GTATTTGGCC ATTAAGAGAC TTAGAAGACT TAAGCCATAG ATTGCTCAGT
50401 GAGACCCCGA GGGCAAATGG TCTGAAGGTG AATAGATCAT TTCACCTTTA AGAGAGCAGG
50461 TAGGAAGCTA TAAATCCAAG ATTAATAAGT TGACTGAAC TTTAAAGAAG AAACCTAAT
50521 CTTGAGCCAC CCTATCCTTG CTCCACCTTC TGCTGCAAGC AAACAGAAAT GCTGAAATTC
50581 AACACTCACA AAGGCTGGTA AGCTGGAAT GACAAAAATT ACTCCTGGGA AAGTCAGATT
50641 TAGAATTAGG CCATATTTGT TGGGTTTCAG ATTTTCATGT AACTTGGGA AAGGGTTTAG
50701 CTTATAGGCA CATGCATGAA GGGAACTGGT ATAGGGCTGT GTTCATAAGG TCAAGAGTTG
50761 AAGGCCAGGC ATGGAGGCTC TTGCCTGTAA TCCAGCACT TGGGAGGCC GAGGCAGGAG
50821 GATGGCTTGA GCCCAGGAAT TCAAGACCAG CCTGGGAAAC ATAGGGAGAT GCTGTCTTCA
50881 CAAAACAATT AAAAAATAAA ATTAGTCAGG TGTGGTGGCA CACACTTGTG GTCCCAGCCA
50941 CTCAGGAGGT TGGGAAGATC ACTTAAGCCT GGGACATTGA GGCTGTAGTC AGCCATGATA
51001 GTGCTACTGC ACACCACTCT AGGTGACAGA ATGAGACCCT GTCTCCAAA AAAGAGCTGT
51061 ATCCACATCC CAGGAAAGTG GTTGAAGATC TACTTTTCTC TGTAACCTA ATAAAGAATA
51121 GAGTGACAAA TGTGTGTTGT GGAAGAAAT GGGGTGAGAG CTACGTAGAT GCAAAACAAT
51181 ACATCCCCAC ATACCACTTG TTAATCATCC TTTTCCACCC ACTTATGGGA TGAATTGCAT
51241 CTCCCAAAA GATACTCTGT CCTAACCTC AGTACCTGTG AACCTGACCT TATCTGGAAT
51301 ACGGTGAGTT CACTGGTTAA GAAGAGATTA TAGTGAATA GGGTGAGTCC TCCAACCAAT
51361 GACTGGGGTC CTCACAGACA CAGAGGGATG ATGGCCAGGT AGAGATGGAG GCAGAGATTG
51421 GAGTTATGCT GCCACAAACC AAACACAGGA AGCTGCTAGA AGTGGAACA GGCAAGAAAG
51481 AATCCTTCCC CAGAGGCTAC AGAGGGATCT TGGCCCTGAT AATACCTTGA TCTCACTGG
51541 CCTACGTAAC TGTGAGAGAA TAAATTTCTT TTGTTCTAAG CCACCCAGTT GATAGTACTT
51601 TGTTACGGCA GCCCTAAGGA ACTTGATATA CATTCTTTTT ACTGTCATAG AAGTTTTGAA
51661 TCTTTTAAAGT AGGTCTGTAC CCTTCCTCCC AGTGTCAACG CATGGAATTC CTCTCCTTGT
51721 GCCTTGAAAA GTGAAAGGTG TTTGAAGTGG TAATGAAAGA AATCTCAGCA TGAGGCCAGA
51781 TGCTGTACCT CACACCTGTA ATCTCAGCAC TTCGGGAGGA TGAGGCCGGC AGATCACTTG

Figure 9 (Page 16 of 74)

SUBSTITUTE SHEET (RULE 26)

105/162

51841 AGGTCAGGAG TTCTAGACTA CTCTGGCCAA CATGGTGAAA CCCCATCTCT ACTAAAAACA
51901 AAAAATGTTA TCCTAGCCGG GCATGGTGCC TGTAGTCCCA GCTACTCAGG AGGCTGAGGC
51961 AGGAGAATTG CTTGAACCCG GGAGGTGGAG GTTGCACTGA ACTGAGATCA CGCCACTGCA
52021 CTCTAGCCTT GGTGAGAGAG CAAGACTTGG TCTTAAAAAA GAGAAAAAGAA AAATGAAATT
52081 TCAGCATTAT AGAATAAAAA TGTTTCCCTT TCCCCCATA CTTTAAAAAA GCAGAAGTCT
52141 GCATCATAAA ATGGTCTTTG CCAATGTTAT TTTTATTATA ACAAAGGAAT CTTGCAAGGC
52201 TACCAGATCT CAGCAATTGT CACTATGTTT TGTAAAAATC ACTTCCTAAA ATGTCTGAAT
52261 TGACTGCTTG TCTCATTAT TTGTTTCTCG TGTCATACTG CAATGGATAT CTGTCTTGTT
52321 AGTATAAATA TTTGTGCATT TTGTTGTTGT TAAAACAGCT TTTTGGCCT GTCTTCTTCC
52381 ACCTATGAGG TAATATAAAA CTCATGTTTA ACACTTATTT TTGTAGGAGG ACAAGCTACA
52441 GACAAAACCC CTCAGACACT GAGTTAAAGA AGGAAGGGCT TTATTCAGCT GGGAGCTTTG
52501 GCAAGACTCA CATCTCCAAA AACCAGAGCTC CCTGAGTGAG CAATTCCTGT CCCTTTTAAG
52561 GGCTTGCAAC TCTAAGGGGG TCTGTGTGAG AGGGTCATGA TCGACTGAGC AAGTGGGGGT
52621 ATGTGACTGG CAGCTGCATG CACCAGTAAT CAGAACAGAA CAGGGATTTT CACAGTGTTC
52681 TTCCATACAA TGTCTGGAAT CTATAGATAA CATAACCGGT TAGGTCGGGG GTCAATCTTT
52741 AACCAGACCC AGGGTGCAAC ACCAGGCTGT CTGCCTGTGG ATTTCAATTC TGCTTTTAG
52801 CTTTTACTTT TCTTTCTTT GGAGGCAAAA ATTGGGCATA AGACAATATG AGGGGTGGTC
52861 GCCTCACTTA TTCACCCCTT TTGAGAATCT CACTCATTAG TGGGAGTTCT CACTTTTATT
52921 CTCCTACCT ATGTCTTCTT GAAAGACAGA TTGATAATGA TTCATATAGT AACTTGTGTC
52981 TGAAGCATT TGGTGAGCTA AGGTAGTGAT GAAGCTTTTT ATCATTGGGA GAAGTACAGG
53041 TAGCAAACAA GGAAGCAGTA AGCAGGTTTC TATTAATATT ATAACCTCTA TTATAAGAGT
53101 TTTAAATCTT CTTAGCACTC GGAACCATTT TTCAAACATG GCCCCAGAAA CAAATCCATA
53161 CCACACCTAC ATGGGCACAT GTGCCACTTT TGTCAATTT CTAACATGT CTTCACTAC
53221 TTGCCCTTAA TCATCTATGT GTAGACAGCA ATTAGTAAGG TTAAATTTCC TACAGACCCC
53281 TCCTTCAGTT GCTAGCAAGT AGTCGAGAGC CAATCCATTT TGATAGATAG CATTTTGCAT
53341 CTGAGTTTCT TGCCAGGCCA CAGTAGTCAG GGCTCTGCTG GTCTTATTAG TAATTATTTT
53401 TAAGACAGCT TGTAAACGTA TGATTAGCTT GAGCATGTAA ATGGGGGTCC CATATCCCCA
53461 CAAGCCGTCT TGTGCCCAAG TAGCAGGCCC ATAATATTGT ATGATTCTCT CAGGGGGCCA
53521 TTCATTATTT TTCCAATTTT CTATAGCTAT GCTTTTTTTT TTTTTTTTTT TTTTTTTTTT
53581 TTGCGGGAAG CATATACAGG GAAGCCAGG AGTTTGCTG TCTTTATGGG CAGTAGGAAG
53641 AAAGATGGTT TAATAGTGTC AATAACACAA CTACCTGCCC ACTGGTCAGG TAATTTGGCA
53701 TAAGCTGTAT GCCACATAT CCAGTATAAT CCAGTGGGGG CTGTCCAGTC CCGGTGGGAC
53761 TCTGGGTGGG TCCACACAGT TTGCAACTTT GGAATTTTAC TAAATAGATT TTTCTTAGTG
53821 TGGTTTGAAC TCCACTAGGT GGCTGTTTTT ATAGTACTAT TATACAGTTT TTGCCAAGG
53881 CAGCTGAGTC TTCCACAGG AAGGGTGAAG TCCTTCCCCA CTTTGTCTAT ACAGTATTGT
53941 CTAATGATTG AGGCTTTTAG GACCCAGAAG TTATCAGGGT GAGTCTTTTG AGCTGGGAAT
54001 TTATCAGGAA CTGGGTCTGT AGGTACTAAT TCTCGTGCTT CCCATGGCCA TTGATCTCCC
54061 ATTACAGTTC CTCCACATAC ATACATAACA TGAAGTGACA TTGAGAGACT GGGCTACATG
54121 CTCAGCTAAT TGCAAAAAACA AATTTCTTGT TTTTCTGGA ATTTCTAGTA CTGGCACATT
54181 CAGTTCATCA TAAGAAGGTT TGAAATACTG GCTCAGGGGA GCATTTATAA ACTTCTCCTC
54241 AAACCACCAT ATTTACTCAA GGATCCAGTC CAGCCCCAAC TATTTCTAAG GTTACAGAT
54301 CCCCTTTTTT CCAGTGAGAA TCAAGGGGGT TGGTTATTAC TAGTTCTAAG GGGTTACACT
54361 GACCACTGGT ACAGGAAGGG CCACTTTTCC CTTTCTGAAG GTGGACAGGA TTCTTTTAT
54421 TTTTAAACCA AGTTGCCTAA ATGACACAAG ACCAGTATCT ACATTTATTT CCACGCAGTC
54481 TTAATTCATG ACAAGCGTAC TTATTTTCTG CCATATAGCC TCTTCTCTAA TGAACAGAAC
54541 CACATCCTAT TTCTAACTTA TTACTATTAA TGACAGCACA GGCATCAAAT TTCAAGGTGA
54601 CTTGTTTGGG CATTCCTTTT TCTTCTGTTT TGGCTAACAC TTTACTCGTA TCGTTTATGA
54661 ACCCCACCA GTCTCAGTC CTCAATCTTA TTTCAAAAAC TGTGGTCTG GAGGCTCAG
54721 ATGGGTCATA ACACACATCA GGTGGTTCAT TTCTGGGCT ACCTGCCTTG TATAGAATAG
54781 CATTATACAA ACAAGTTATT TTTAGAGTCT TTGTACACTT ATAATAACCA TAAAATAATA
54841 AGACTGTAGC AACTTTTTGT CCTACCTCAG TGACTTGATG TATACACTGG GAACAGCCCT
54901 CAGTCTGAGG AAGGTTAGT GAAGTCTTTA CTGTGCAAGT CCAAATTTTA AGGAAAATGA
54961 GTCCCTTGAT GAGTTTTCTC ATGTTTCGGC CATGCATGGA CCAGTCAGCT TCCGGGTGTG
55021 ACTGGAGCAG GGCTTGTGT CTTCTTCAGT CACTTTGCAG GCGTTGGCGA AGCTGCCACG

Figure 9 (Page 17 of 74)

SUBSTITUTE SHEET (RULE 26)

106/162

55081 TACAGCTCAC AGTCTACTGA TGTTCAAGGA TGGTCTTGGA AGTTGGGCCC ACTAGAATTA
55141 ACTGAGTCCA ATACCTCTAC TCAGTCACTT TCAACTGGGC TTTCTGATAC CAGGAGCAAG
55201 GTGGCAGGTT TTAGGGTGTG GCAAATTTCA ATGGTTATGC AGGGATTTTC ACATAGCAAA
55261 CTTTGGTACT TGGTTAATCT AGCATTGTG AGCCAATGAT GTATTTATTA AAGTCACCAC
55321 AGCATGGAGG GCCTTTAAGT TTAGGTTTTG TCCAAGAGTT AGCTTATCTG CCTCTTGTC
55381 TAGCAGGGCT GTTGCTGCCA AGGCTCTTAA GCATGGAGGC CAACCCTTAG AAACCTCCATC
55441 TAGTTGTTTTG GAGGCCCAGC CTCGGCCAGG GCCCCACAGT CTGGGTCAAA ACTCCAACCG
55501 CCATTTTTTC TCTTTCTGAC ACATAGAGTG TAAAGGGTTT TGTCAGGTCA GGTAGCCCCA
55561 GGGCTGGGGC CGACATGAGT TTTTCTTTTA ACTCATGAAA AACTCATTGC TGTTGGTTGT
55621 AATAGATGTA GTTTATCCAA TCTACATTTT TATTAACGTG CACCCACCAA AATATTGACT
55681 CAAATCCTGC AGCTATTTGA TTTTGGGATT TAAATTGATC TGCTATTCCC TGTGGGACTC
55741 CAATTGCATC TAAATAGATG TGAGAGTTGA AAGACACATA AGGGTCTTCT CTTGCTTTAC
55801 GATGTCTTAT TTTTCTCCC TCTGGTTGAT GAAATGCTAG GGTGAAAGGG ATAGCCAACT
55861 GGAATAAGT ACAAGTGCCG CTCCAGTTAT TGGCAGAGT GCCCAGTAAA GGTCCACCAC
55921 AATACCACCA CACATCCGCT TGGGGATGAA CAAAGGCTGA CTGATTGAGA AGCTCCTGAA
55981 AATTCTTAAG CTCACTGCAT CCCTTCAGGT CTCCAAGGAA TGCTAAGTTT CCTCCCTGTC
56041 ATGAGAGACA AGAAGTGAAC TTAGTTTTG GAGATGGAAG CTGGATGGCC CTCAGGGGTT
56101 GACCTGCAGG GTGCTGGACT TGGGATATA GCAGAGAGAG CTTGGCACGA CTTATTACTC
56161 CAGGCTGTAG CATCCTGGAA AACAGTTACC ATGCAGCCCA TGCCTGGTCA ACAGGAGGAC
56221 CACCTTAGTG GAAAGGGGAT AATCTGGCCC TCTGGCCTGC CATGTGCACA AGCATAAACA
56281 TTGGTTTTGT TTAATGTGTG GACAGAATAT TTGATCCATT CCAACTGGGC ATTTGCATCT
56341 TGGTATCCTG CTTAATTATC AAAGTTTGT TTAAGTCTTT AACTTCTATG ACCCTCTAGT
56401 AAAATGAATG TATGATTTTA GGAAATTACA AAAACCGGTT GGGGCAGTCC ATCCTCGCTC
56461 TTTAGTGGTC CACACAACAT TCGACCAACT ATGGCATAAA AGCTCTACAT CAGGGGGCAA
56521 GACTCCTCGT TGACACTGGG GTCTTTATTG AAATCTCTCT GGATTAAATG GTCTCAGTTT
56581 ACTAAGGCTC AGTCTGAGGA GAGTCAGGAG GGACAGAGGT ACTTTTCTGA AGTACAGAGA
56641 TGTCTTCGAC TTGGCAAGTC CCCACAGGAT ATAACAAGGC AAGCATTAAA TTCAATAGTT
56701 TGAGGCAAAA TTGACTTGGT TATGTTAATA ACTAGATGGT CAGAAATAGA GTGAGGGAAG
56761 AAGAAAGAGT AATAGAATAG ATGAAGGAGT TAAATTTTTC TTAGCTTTAG TTTGGTAGGG
56821 TTTTCCCCTG GGAATATGGC CCATGACTCT GGAGGGGGTG GCACTTTCTT GACTCGGGTG
56881 TGATGAGTCC ATCCCTTTT CACCGTATGA ACAACAGTCT CGGTGGTTAG CAGCACAAGG
56941 TAGGGTCCTT CCTAGGCTGG CTCAAGTTTT CTTCTTTCC ACCCTTTGAT GAGAACATGA
57001 TCTTCAGGCT GGTGCTGGTT TACAGAAAAT TCTAGGGGTG GTACATGTGC TAAAAGACTT
57061 TTAGTTTTGA GGGAAAGGAA AGTGGAAGAT AAACCAAGTA TATACTTTT AAGAAGTTGA
57121 CCTTTTGTG TAAATGTGGG GACATCAGCA GTGGACTTTA TAGTCCTTGG TGCCTTCTTA
57181 CTGAGAAATT TCCTTTAGCA CCTATTTTTA TTAGTTTTTA GACCAAAGAA AGTCAAATGC
57241 CATTTTATAT TTGACAACGC TTCTGTATG TTTATACCAG ATAAGCTAGA TTTCACTTTT
57301 ATATTGGTGT GTTATTAATG TTAACCTTAG TTTAATAAAA ACTCTGTAGA CATATTTATT
57361 TGATTTTTTA TGTCTGACCA TAAGGTAAGA TTTTATAGA CTTTTCTTTA ACCTTTTATA
57421 ATTTTTGTGA AAGAACAGGT TAGTGCTTTA AGAAAAACCC GTTGTGTTTT TATTTTAATG
57481 TTCAGTTCAC AGAAAACTG TATGATACCC CTTAACTTTA GCCAATATGT TTAGACACAG
57541 AATTTTCTTT ACAATTAAGG TTTCAAACT TGCTTAAACC TTCAAAACAA TTTTGTAAAC
57601 CTTTTAATGT AGGTAAAAAT CCACATTCTT ATGCATCCTC ATAATCCTTT TACCAAAGGT
57661 ATATTTTACT TTCCTTACAT ACCTTGACCA TAACTGTTT ATTCAATAGT TTTACATTTA
57721 GAAGGAGGCC TAATTACTTT TAAATTATAC AACATTTCTT GCATAAATTT ATTTTCTAA
57781 CACACATTTT TTTCTAGACT TTCACAGACA ATTCTTCGAC ATGCCTCAAC TTTCTGACTT
57841 ATTGCAACA TCCCTTTCTT TAAACAATA GTTAATTTAT CTCAGGACAA GGATTTTCCA
57901 TACAACATC TTTTTTATAT AAATTCTGCC TCCTCTTTAT TTCCTTTTTT TTTTCCGAG
57961 GATGATAACC ATTCTTTTCC AAAGCGAACT TCTTTTATGT CTGTGGACTA GACTGTCTAA
58021 GGCCACAAGA TTAGAAGTTA CTATAATACA TGTTACACTG TTAACTTTTA GCAAACCTTA
58081 CTTTTGTGA AAACCTTGTA AGTTGGGAT TTCAATTATC CTTTGCTATT AATAAGACCT
58141 TATTTAGTCC AAATTAACCT AGAATTGGTA TAGATGGCTT TTTTTTTTTT TTTAATTACC
58201 TGGGAGGAAC CATCTATCCT CCGTCTCTGA AGGGAGTCC TCCTAGGTCT GGTCAGAGCT
58261 TTGTATGGTA ATTAAGATTT AGATCCCCTG TTAGGAAACC TGCCGGGTTA AGAGAATTTT

Figure 9 (Page 18 of 74)

107/162

58321 CAGTGGTTAA TGTTAAATCA TCTTCTTTTT TCTTTTTTCC TTAGGATACT TCTGAACCGG
58381 TGAGGTGTGC TCACAATGAG GTTTCCTGTA AAAGTTATTT TTTTACTTTC TTCTGTTAGC
58441 AAAGCAGTTG CCGCTACAGA TTGAATGCAT TTGGGCCATC CGCGGGTTAC TGGGTTAAGG
58501 ATTTTTGATA GGAAGGCCTT AATGCTTTTG GAATATGCCC TGACAACAAA GTGCCAGTTC
58561 CTTCCCGGTG TTCAGCCACT GCGTTGATCC TCCACGAGGG CCTGCCACGT GCTGCTCTGG
58621 TGAGGCGTTC CACCGGGGCA ATTGCCTACC TGGGAGCGCT CTCCAGATCT GTGTGCTCA
58681 AACTGGCTGG AGTTCCCCGT AGGGATGCTC CACAGGGCAG GCCTAAGTCG CCTAAGGGGC
58741 TGCCTTGACC GTCCGTAAAT CACCTCTGTC TCCAAAAACC AGCTCCCTGA GTGAGCAATT
58801 CCTGTCCCTT TTAAGGGCTT ACAACTCTAA GGGGGTCTGC ATGAGAGGGT CGTGATTGAT
58861 TGAGCAAGCA GGGGGTACGT GACTGGGGCT GCATGCATCA GTAATCAGAA CAGAACAGAA
58921 CAGCACAGGG ATTTTCACAA TGCTTTTCCA TACAATGTCT GGAATCTATA GATAACATAA
58981 CCTGTTAGGT CAAAGGTCGA TCTTTAACCA GACCCAGGGT GCGGTGCCGG GCTGTTTGCC
59041 TGTGGATTTC ATTTCTCCCT TTAAATTTTT ACTTTTTCTT TCTTTGGAGG CAGAAATTGG
59101 GCATAAGACA ATATGAGGGG TGGTCTCCTC CCTTAATTTA AACAAAATTT TCAAAGTCCCT
59161 ACCCCAAGTA AATTGGCAAA TATTAATAAA GTTATGGCAT AGAAAATAAA AATGATTGTA
59221 AAAGGCGTAA AGATATTTCT GTGGGGAAAA CATTGTTC TAAGTTATCA GTTAAATTC
59281 TGTGAAAAAT AACCACTAGA GACCCTAAAG TACCAGGGG CTAATAATAA GAAGGGAGGA
59341 ACACCTCTC AGTCCCCACC GTTACCTCCC CAGAAGGGAA GAGGAAGAGG GTGACTCCAG
59401 GAGAGCTGTG GTCTCCCTC CCCATATGTC CACATATACC TGACCTCCCC TCCCCAAAT
59461 ATATACCCAA TATCTCTCCC ATATATACAT ATTTATCTGA CCTCTCCACA TATGTATACC
59521 TAACTTTCT CTATATATCC ACATATACCT AACCTCTCA CACACATATA GCTGACCTCC
59581 AGTGGAGGAA AATGGGGAAG AGAGAAGAAG TTATCAAAGG ATAAATCTAG GTCATACTCA
59641 GAAATGTGAA AAACAAAAAC CACACACAGA AAAAAAAAC ACACACAAA AAGAAAATGA
59701 TAAATTTGTT TGTGTCAAAA TTAAGAATTC CGGTTCAATG AAGGATCCCA TGGATAAAGT
59761 TAAGACACTG CTGTAAGGAT GGTAGAGAAT TAAATGTCTG AATCAGACGA AAGGATGAGT
59821 AATTAGAATG CACAAGGCCA AGAAGAACAA AACAGAACT CCACATAAAA AATGTATGAG
59881 GCCGGGCGCG GTGGCTCATG CCAGTAATCC CAGCGCTTTG GGAGGCCAGG GCGGGCCGAT
59941 CAGGAGTTTG AGACCAGGCT GGCCAACATT GTGAAACCCC ATCTCTACAA AAAATACAAA
60001 AAATTAGCCG GCGGTGGTGG TGGGTGCCTA TAATCCCAGC TACTTGGGAG GCTGAGGCAG
60061 GAGAATCACT TAACTCAGG AGGCAGAGGT TGCAGTGAGC TGAGATCACA CCATTGCACT
60121 CCAGCCTGGG TGACAGTGTG AGACTCTGTC TCAAAAAAAA AAAAAAATTA TATATATATA
60181 TATATATATA TATATATATA TATATATATA TGAAATAAAT GAACAAGAAA TTTAGATACA
60241 GGAAATCCA AAGCACTTGG TAATGAAAGA AAGGTAAAGT GATGTGTCCT TTTGCATTTA
60301 AAAGAGAGCA TTAACAAATT AGAGAGCTGA ATAATGCTCA GTATTGGTGT GGATATGGAG
60361 ACTCAGGAAT CCTCATACAC TGCTGATGGG AGTGCCCACT CCCTGGGAAT ATTTTCCAAA
60421 TATCATCTCA AACATATCCC ATAAAGGTGA CAGGAAAGTG TGGGCTGACT GATATCCTTC
60481 ACTGAGAGAG GTGGAGGTAA AATGAAGTCA CTGCACAATA TAGAGTTGGA AGCAATGGAT
60541 TAGATGTCCA CATAGTTACG TGGAGAATC CGTAAGATAC ACACACACAC ACACACACAC
60601 ACCTTTGTGT ATATTGTTCC TGGCAGGTAG GCATGGAGGT TTAGAGGCTT TCTACATCAC
60661 ACCTACTGCA CACAGTAAAT GGCCAGGCTG AGCACTGACT TCCATGAAGG GAGATTGAAG
60721 GTAAGAGATT GAAGATTGTT CCCTGGTCTG GGACCCTGCA ACTGAATATG CAGAAAAAAG
60781 TACACCCCGC CACCCCGCTT CCCATCTTTC CTACCTGATT AGAATAGCTT TTTCAGAAAA
60841 CGTTGGCCAG GGGTTGTGGC TCACACCTGT AATCCCAGCA CTTTGGGAGG CTGAGGCGGG
60901 CAGATCATCT GAGGTCAGAA GTTCCAGACC AGCCTGGCCA ACATGGCGAA ACCCATCTC
60961 TACTAAAAAT ATAAAAAATT AGCAGGGCAT GGTGGCACAC ACCTGTCATC CCAGCTACTC
61021 GGGAGCCTGA GGCAGGAGAC TCACTTGAAG CACAGTGATG GAGGTTGAAG TTAGCTGAGA
61081 TCTTGCCACT GCACTCCAGC CTGGACAACA GAGTGACACT TTGTCTCAAC AACAACAACA
61141 AAACCCACCA AAACTTTAAA TCTACCTATG GCCAAATGCC TGCTAAAATG AGCACCCAAAG
61201 AAGCAGTGT CAGGAAAGTC AGATGAATAC CCTAAAATTA GATGCAATGT TGGTGGTCA
61261 CAGTGGCTCA GGCCCTGTAA TCCCAATCCT TCTTGGGAGG CCGAGGCGAC AGATCGCTTA
61321 AGCTCAGGAG ATCGAGACCA GTCTGGACAA CATGGTGAGA CCGTGTCTCT AAAAAACGT
61381 AAAAAATGA GCTGGGAGTG GTGGCGCACA CCTGTAGTCC CAGCTACTCA GGAAGCTGAG
61441 GTGGGAGGAT CTCTGAACC CAGAAGGCGG AGACTGCAGT GAGCAGAGAT CATGCCACTA
61501 CACCCAGCC TGGATGATAG AGCCAGACCC CCATCTCCAG AAAAAAAT AAAGAGAGAG

Figure 9 (Page 19 of 74)

SUBSTITUTE SHEET (RULE 26)

108/162

61561 AGAGATGCAA TATTTAGGGT TCAACAAGAC TGAACCTCTG ACTCCTTTCC CTACCTCTCC
61621 AGCATGTTAG ATTCTGGGTC CTTTCATCCTA ACCCCTGTG CATGCCATAG CCACCCTGTG
61681 GTACCAACTT TGGGAAGCCTG GATCTTCATC CCCTCATGAT AATGAGTGTC CCATTTCAGGT
61741 CTCCATGCTC AGCTTGGCAA GAGTATCTGT CTTCTCCTCA TGGGACGGTC ACATTACCCC
61801 AGCACTGACA GGTTCCATT CCACTAGGGT GGCACCCTAT ATGGTCTGAG TCCAGGCCTT
61861 CCTGGTCCCT CAGTAATCTC AGCATGGTAG CACAATCGAA AAGGGCTAGG CACGGCAGCA
61921 CCAATTTCCCA CCAAGAGGTC TGATGGCTCA TCACATAGAC TGAAGGAGAT TCTGAAGAGC
61981 AGAGGTGGAA TGAAGAATGA ATCCTGGGCT CTGCTCTTCC TAGGCCTGTC TTCCTCTCTC
62041 CCGAGATGTT AGCTAACTCA TGAGAGCCAG AAACCAACTG CAGGCTGGCC TCAGGCACTT
62101 AGGTAGTGCT TCAGCCTCAG CAGTCCACAT TCTAGGAACC CTCATAATAT GGGTTGAAGT
62161 ATGCATTCCC AAAAAAATAA AGTTGTTGAA GTCCTAACCA CCACTACTGA AATGGGAAAA
62221 GTTCCCTTGT CCCGCTCGCA TGGCATGTGA TAGGAGTGTG GCTAATTTCT TCAGTGCCTG
62281 GCTGCTCAAA CCTCTAGGGG AACAGTAAGA CGGGCAGGTT GTGGGTCTCC AACCCTATGA
62341 CCCCACCACA GTGTCTAGGG TTGAATGTTT ACAGCTCCTG AAGCCACAGT GGGTGTGTGT
62401 TACAGGGTGC TCTTTTAGTT TTGCCATTTA TAGGCAGCTG GTGTTAACCA ACTCAATTAG
62461 ACCGTCTACC TTGTCCCAAG GACAGAAGAA GGCTTTCTGT ATCCCAGGTT CTGCTCTTGG
62521 TGTACCGGAA TAAATCAGAC CACACCTGGG CTTAGAGAAA GAGTGCAAGG TTTTATTAAG
62581 TGGAGGTAGC TCTCAGCAGT TGGGCAAAGC CAAAAGTGGA TGGAGTGGGA AAGTTTTCCC
62641 TTGGAGTCAG CCACTCAGTG GCCCAGGCTC TCCTGCAACC ACCCCAGTCA AATTCCGCTT
62701 CATTTTGCCA GGCAAACGTT TGTTGTGTGC TCTTCTGCCA GTGTGCTCCC CTGGACGTCC
62761 AGCTATTCTG GTCTTGTGGC AGGCTAGGGG AGGTCTTGGG AAATGCAACA TTTGGGCAGG
62821 AAAACAAAAA TGCCTGTCTT CACCGTGGTC CCTGGGCACA GGCCTGGGGG TGGAGCCCTA
62881 GCCGGGGACC ACGCCCTTCC CTTCCCTACT TCCATATCAT TTAAGGGGAC CATGCCCTTC
62941 CCTTCCAGC ACTTTCCCTT TCCTGTATCA GGACCTGTGA ATGTGGCCTT ATTTGGAAAT
63001 AGGGTCTTTG CACTTCATCA GTTAAGATAA GAGTGGGCTC TAACCCAAACA TAAAGGGTGT
63061 CTTTATAAAA AGGAGAAATG TCATACACAG AGACTGACAC CTATAGAGAG AAAATGTGGT
63121 GAGTAGACAC AGGGAGAATC ACCATTCAAG TCAAGCAATG AGTCTGGGGA TACCAGAAGC
63181 TGGGAGAGAA ACCTGGAACA GATTATCCCT CATTGCCTTC AGAAGGAATC AAACCTGATG
63241 ATACTTTGAT TTCAGACTTC CAGCTTCCAG GACTGTGTGA CGATAAATAT CTGTTGTATA
63301 GCCAACAAAGT TTGAGGTACT TTGTTACTGC AGCCCCAGAA AACTAATACA GTAGGTACTA
63361 TGGACTGAAT TGTGACTCCC CGTCGCAAAA TTCATATGTT GAAACCCTAA CCCCAGTGT
63421 GATGGTACTT GGAGCTGGGG CGTTTGGGAA GTCATTATAT TTAGACAAAC TCATCAGGAT
63481 GTGTCTCTCA TGATGAAATT CATGCCCTTA TTAAGAGAGA CAACAGGCCA GGTGAGTGG
63541 CTCATGCCTG TAATCCAGC ACTTTGGGAG GCTGAGGTGG ATGGATCACC TGAGGTTGGG
63601 AGTTTGAGAC CAGCCTGGCC AACATGGTAA AACCCCATGT CTAATAAAA TACAAAAATT
63661 GGCCAGGTGT GGTGGTGAC GCTTGACTC CCAGCTACTT GGGAGGCTGA GGCAGGAGAA
63721 TCCCTTGAAC CCAGGAGGTG GAAGTTGCAG TGAGATCACA CCACTGTACT CTAGCCTGGG
63781 TGATAGAGAC TCCATCTCAA AAAAAAAGAC AAAAAAAGAC AATAGAGCCA GGTGCTGCAG
63841 CTGATGCCTG TAATTCCAAC ACTATGAGAG GCTGAAGCAG GAGGCTCGCT TTAGCCAGG
63901 AGTTCAAGAC CAGCTTGGAC AAAATAGTGA GACCCCAAC TTCTAAAAAT TTAAGAAATG
63961 AACTGGGTGT GGTGGTACAC ATCTGAGGCT CCAGCTACTC TGGAGGCTGA GGTGGGAGGA
64021 TTGCTTGAGC CCAGGAGGAG GCTGCAGTGA GCCATTGCTG TCCAGCCTGG GCTACACGAG
64081 AACCTGTCTC GGGAAAAGGA GAAAACAGTG AGACCTCTT TTCTCTCTC CTCTCTCCA
64141 CTGCCTAAGC CCTACAAGCA CAAAAGGAC ACCACATGAG CACATAGTGA GAATGCTGCT
64201 GCCACCAACA AGTCAGGAAG AGAGCGTTCA CCTAGAACT GAATTGGCCA GCACCTGGAT
64261 CTTGGACTTC TGAGCTTCCA GAAGTGTGAG AAAGTTATTT TTTTITTAGC GACTAAGTCT
64321 ATAGTATTTT ATTACAGCAG CTCAAGGTAA CTAACATAGT AGAAGGGATG AATTATGGAG
64381 ATCACAAGTC CACGCCTCCA GAAAAAGACT TCCCTAAAAA TTAGTCTGAG CAAAATTCGA
64441 ATGATGAATT ATTTTAAAGA ACTTTTAAAG GATCTGACAA GTTGCAAGA GCTAGAGAAT
64501 GCTTTACAAC GTGATAATAG AATGCTCTGT GATGACAGAA ATCTTCCAC ACTGTTCAAA
64561 ACTAGCTACT GGCCACTTGT GACTATTGTG CACTTGAAT GTGACTGGTG TCTGAGGAGC
64621 AGAATGTTTA ATTTTACTTA ATTTTAAATC ATTACAATAG CTACATGTAG CTAGGGGCTA
64681 CTGGATTGAA CAGCACAGCT CGAGTCTTTT AGAGGGAGAC AGGACTCACC AAGGTGGATG
64741 CTGGTGGCCA AGCAGCAATG GCAGGTAGTA CACACACAAG AGGCAGATGA TACAACACAT

Figure 9 (Page 20 of 74)

SUBSTITUTE SHEET (RULE 26)

109/162

64801 CCTTCCCAA CCTGGAGATA AGCTCACCCC ACAATCCCGC CGCTGAAATA GAGTTGATGT
64861 TACCAATGTG CATTTTATG TCCTTTTCCA TACAGAAAGA TCATTCAACA AGTACTATGG
64921 TACTTAAAA ACAACATTCA ATTCATTATT ATGACAAAAT TAAATTAATA GCTCTTCCTT
64981 AAACCTTTTAA ATTCAATTTA CAATGCTTAC TATTGGCATT TATTAATCTA CCAATTTTTT
65041 CCCATAGAAC CCATAGAAC AATAATCTAC CAAATTTTTA ACATTCATTT TTGGCAAGGC
65101 TTTTGCAATT TGACGAACCT TAAGAAGAAA ACTTATAAAT TGCAATTTTT AAATCTGACA
65161 TACTGGACTT TTAAGTATC CAATTGACTA ATGAACAAAA CTGCTCCAAA TTTTCAATT
65221 CTTAAAAATC TTAAGACAAT ACTTAATATG GCAAATCTTA ACTTCTTAAA CTTTGTAAGA
65281 ATGCTAATCA ACTTAGATTG GTATAAAGTT GAGTTAAAA TCACAGGATA CATCATCTCA
65341 GCTATAAGTT TTCATGAGTT GAGTTTTTAC AATCACTTGA AATGCTTAGA ATAGGAAATA
65401 CGTATAAATT ATTTAACATA AAATATTGTT ACAAACCTC TGGAGTGTCA GTTCTCTGG
65461 CCAGACTTTA TGCTGCAGCA CCTTGCCTG AGTTCTTGTC CTGCATCCAG GAAGAATTAG
65521 GTACAGAGGC AAGAGTCAAG AAGATTAGTT TTCCAATAGT TCAGCTCACC TAGTTAACTC
65581 CTGTTCACAA TCTTCAAAGT TATCAGAAAC CTGCAATTGA GGGTTATAAT CCATTCTTTG
65641 CAGAGTTTCA AAACAAGACA ACATTGTCT ATGAATGTTA AAATGTCCTA GGGTAGTACAC
65701 AGTCAAAAC ACAATTGACA AAGAAATTTA GTCACCTCTG TGATTTACAA TAGCCTAACA
65761 CAATAACTCT AATTATAACT GATGACACAA ACTCAGATAT CAGAACTCTA GAAATCCCCT
65821 ATAAATTTGG AACACATATT CACAGTTTTT ACTGAAATAT GACCTGAAGA TCAAATATCA
65881 CCTTATTTCA ACAATCCTAT ATAATAAAC GTGTCAAATG ATCCTGTTTA CCTCTCCTTT
65941 GGATACTCCA GGGGCCCTCT GTAGCATCCA AAAGTTAGGG GTTAGCAAAG ACAATTTTGA
66001 AGCTGTAAAG GCTCAAAACA CTTAATGAAC CTCTAGTCAT ATCTGTTCTC TACTCACTAA
66061 ATGCTAGTAG CACCTCTCAG TTGTGGCTAA GCTGGGAGGA TCTCTTGAGC CTAGAAGTTT
66121 GGGGACGCAG TGAGCTATGA TTATGCCACT GCACTCCAGC CTGGGCAACA ATGCAAAATC
66181 CTGTCTCAA AACAATAACA AAAACAAAT TGCCTATGCT GTGGTTATCT CACAATTAAT
66241 AAAAAGGAAA AAAAAGTAT GCAGTCTTTG TAGGTCTTG GGGTTTGTG GAACCTAGAA
66301 AACAAATCCC CAAAATAAAG ACCGCAGAAG CCAAAGTTTT TCTCTGATCT TCTCTGCCC
66361 TCCTGTCTCT GAGTCCCATT CTCCCCGGAG TCTAGCCATA GAAATGAGAA TTCCTCTTCC
66421 TCAAGTTAGG TCATAGAAAT CAAAACACCT TTTCCCCAGA GCCCAGCCAT AAAACCTAAA
66481 AATATTACT TAACCTTTCC TCTGTTTTT TGTGTAATA CTGGCCATA AGAAATTATC
66541 TGAACCTTCT TATTTGATCA TAGATCACC GACCGCATTC CAGAGAGGAT CCAGAAGGAA
66601 GGAATGCTGC ACAGAGAGGC CAAGAAGAAT CTAGACAGAC AGGCCTTGCT GGGTTTCCCT
66661 ACTCTGTTTA TTAGCAATCC TATTTCTACA CGGCGGCCCA TACTTTGTTG AATCTAAAAA
66721 ATAAAAATGG ACAATTTCCC CTGTACATGT TAATACACAT TAATAAATTG GATATAAATT
66781 GGATAATTTA TTAATATACA CATTAAATAA TTGGATGCAG CCGGGTGCAA TGGCTCACGC
66841 CTGTAATCCC AGCACTTTGG GAGCTGAGGC GGGCAGACCA CGAGGTCAAG ACCACCCTAG
66901 CCGAAATGGT GAAACCCCGT CTCTATTAAA AATACAAAAG TTAGCTGGGC GTGGTGGCAC
66961 ATGCCTGTAG TCCCAGCTAC TGGGAGGCT GAGGCAGGAG AATTGCTTGA ACTCGGGAGG
67021 CGGAGGTTGC AGTGAGCCGA GATTGCGCCA CTGCACTCCA GCCTGGTGAC AGAGTGAGAC
67081 TCCGTCTAAA AATAATAATA ATAATAATA TAATAATAAT AATAATAATA ATAAATTGGA
67141 TGCATTTTAT CCTATTAATC TTCCTCTTGT CGGTGGTTTT CAGCGACTCT TCAGAGGCCA
67201 AAGAGTAAGT TTTCCCTTAG CCCCTACAGG TTCTTATGTT TAATTGTTA CTCTCATTTA
67261 AGACATAATT AAAGTGGCTT CTCCATGAAG ATTATTTCTG CATCCATTAT TTGGTAAGAT
67321 TGGCCGTTTT CTCCTTTGAT CTCTACTTCA CACTGACCCA CATAAAACAT CACTGCCTGT
67381 TTTTTGTTG TTGTTGTTG GAGACGGAGT CTGCTCTGT TGCCAGGCT GGAGTGCAGT
67441 GGTGTGATCT CCGCTCACTG CAAGCTCCGC CTCCCGGATT CACGCCATTC TCCTGCCTCA
67501 GCCTCCTGAG CAGCTGGGAC TACAGGCACC CACCACCAAG CCCGGCTAAT TTTTGTATTT
67561 TTAGTAGATA CGGGGTTTCA CTTTGTAAAC CAGGATGGTC TCGATCTCCT GACCTCGTGA
67621 TCGGCCCCGCC TCAGCCTCCC AAAGTGCTGG GATTACAGGA GTGAGCCACT GCGCCCGGCC
67681 CCGTTTTTTT TTTTTGGTT TTTGCATGTC TTCTCCCTTT TACTGTAAAC TATTTCCACT
67741 ACCAGCGTAG TTATCAATTC TACTGCTTAA TAATTGTTTT GGGGAAGTGA ATGCATCAAC
67801 CCACATGAAT TTCTTGCTA TTTGACAATT TATTCTCTT AGGAATAGTA TTAACCTCTA
67861 AGGTCTGGG AGCCAGTCTC TGTACTTGGC TGCTCCAGGG TCCTACTTCA GTTTCCTCAGC
67921 TTCTCAGTAC TGTACTGTC AATTGTGGG AATAATTATT TTTGTCCACC AAAAGACTCT
67981 GTATGTGAAT GAGTTTGAA ATCTGCTGAG TAATACAGTG TCAACCCAGT TAATGATTTG

Figure 9 (Page 21 of 74)

SUBSTITUTE SHEET (RULE 26)

110/162

68041 CCGGGCGGCT TGATCAGGGG CTGTCCAACCT ACCGGCATTG TGATTGAG CGTCATCTAG
68101 TGTCTGAAAG CACAAACAAC ATCCTACATT GTAAATGCCT TTGGCTACAG AGATTGAAAC
68161 CAAAGCAAAC CTATGTTTTG AATTGTTATT CTTACAGCAGT TCTGCTAGCC TTGAAAAATC
68221 TAAAAGTTAA AAAAAAGCTT TATATTTTCA TTTCTGCCTA AACTCTTTAA AATTGCTAGT
68281 TGACAATTAG ATATTTTCAA TTTAATGAAA TTTTTTTTAA GTTCACAGAT TAATACACAA
68341 TGGGGGAGGG TTCTTATTCT GTTGGACTTT TACATAACCT CCACTTTAGT GCAGTCTGCT
68401 TTATGGGGTC TTGTTTGAGG TGTGTGTGTG TTTAAGGGAA TGTGGTTTAC AATCAAAATA
68461 TTGGGTTGCT CTTAGGCACA TTGTAAAGTC ACACACCTGT ATTCTTATTG ATACATAATG
68521 ATTAATAACA TTATTATTAC AGCCTGATCA CCATCATTAT TGATATATCT AAATAATGAA
68581 TTTTATAATT TTGCTTCCTG TCAGGCAAGA GCCAATTTCA GTGCTACCAT GTTTGTATAG
68641 CAGTATTTAT GTCTGTCATC CTCAGTCATT TTACTTCACT TGTTCCTTAGC CAAACGGCCG
68701 AGAAGCGATG GTCATTTTAC TTCAAAAATG AAAAGAATTA ATATTTTAC GTTTCCTTA
68761 AAGACCCTAT GTTTAACCTC CACTCCCGGG TAAAATGGTC TAGTCCCTCC TTTTCATATC
68821 ATCTCTGATA TCTTTGACAC AGCCACTATT ACCTACCGTT TTCTAGATCC CTATTCTTCA
68881 AACACCACCA TGAAGGTAGA GCCTGTCTGA ATTATTTTCT TGTCCCGTGA ACTCAGTACA
68941 TTGTTAGGCT TCTTGAAGAT GTTGATCAGT TGTGTTGTTG GTGAATGAAT CAGCTAGCAT
69001 GATTTTTCTA GACCACTGAG ACAAGTGTCT AAGACACTTG TTCCTTCCCA TGTCTTGCC
69061 TGCCTGTGCA ATCCATGCAG TCTCATGGCT TCCCAGTGCC TCAGAATTAT CCCCTGTCAA
69121 ACAGGCATTA TAATTTCTGT CCACTGAAAA GGACAAAAAA CTAAGTGTAT AGCTAGAAGT
69181 TAAAAATTAC CGGCCAGGTA CTGTGGCTCA CTCCTGTTAT TCCAACATTT TGGGAGGCTG
69241 AGGCGGGCAG ATCACCTGAG GTCAGGAATT CGATACCAGG CTGGCTAACA TGGCGACCCC
69301 GTCTCTATCA AAAATGTAAA AGTTAGCCAG GTGTGGTGGC TCGCACCTGT GGCCCCAGCT
69361 ACTCAGGAGG CTGAGGCAGG AGGATCGTTT GAGCCCTGGA GGTGAGGCT GCAGAAAAAT
69421 AGGAATATAC TCTCTTCAA GAGTTCGTGG TTTTGACTGC CACCTAGCGT ACATCAGAAA
69481 AACC GCATGA CATAGGAAAT GCCTGTGACA GAGGGGTAAG GTGAGAGAGG TTGATGAAGA
69541 ATGTATTGAA GGAGTGAAAA CGCTTCCATC CCTTACTTA CTAAATATAT TAGTTAAGTA
69601 GTTGGGGCAT ATTTAATTC ATGCATTTTG TAGATAGAAA AACAAAAGTT TTATTCTGTT
69661 TGATTTAGTT GATACTTTAA TATGTGTGTG TTTAGGATGC ATGATTTATA ATCAGTCTGC
69721 AGCACTTCTT GGAGAAGTCT GAATTCTCAT TCTCCATTTT CTTATTGGCA ACGTGAGAAT
69781 GATTACAATG GTGGTTGTCT CATAGAATGC AGGGAGTCAG AATGAAAATA GTCCATATAA
69841 TGCCTGGTGC AGAGGAAGGG TTCAGTTAAC TGTCTGTATT AATATTACTG ATAACAGTCA
69901 TGACAAACAA AAGCTTAACA ACAACACCAC CAACAACAGT TGCAGAATTG AGCCACCAAT
69961 TTGCACACAA GATTGTAGGT AGGATGTTTT AGAAAAGTTA TTATTTAATA TATGTATATA
70021 TTTTGTACT TAAAATATGT CAGAGGTTGT TCTAAGAACT ATTTAAATGT TAACTCCTTA
70081 ATCCTCATAA TGACCCATGA AACAGGTAGG CTTATTATTG TCTCTTTACA TGTGAGAACA
70141 CTGAGACACG AAAAGGTTTA TTAACCTACC CAAAGTCACA CAGCTGGTAA AACGGCAAAA
70201 TTGAATTTGA ACTCAGACAT TCCAGGTTCC AAGACAGTCT AATTATCTT TTGACTAATA
70261 TACTAAGCTG CCTCTGTATT TTTCTTGAT TACTTTGTAA AAGTATGAGG AAAATATAAG
70321 TGCTTCAAGT AACCATGAAA AATATAAACA ATCTATGTAT CAACTGAAGC ATAATTACAA
70381 ATCCTTTGAT AAGCAAACAT AATAAAAATT TGATATCAAT CAAAACCTTT ATGTAATGTA
70441 AGCAGGTTGA GATGAATTCT ATAGTAAAAA AGTGCAGAGT GCTGGAATAC CATGCTCCTA
70501 ATATATTGGC TAGGCACACC TGCTGTCTAT CAAAGGTATG CACACACCTT GGATACAGAA
70561 AGTTGGGACT GGGTAGTTAT GTGAGTGTC TCAGAATTCT TTCCCACTTG GGAAAGAATT
70621 GTCCATCATA AGCTTGATG ATGGACAAGG AGTGAGCTCC CAGAACAGTG ATGTGGGGAT
70681 ACATCCTCAC ATCACAGTGA GAATGAGTGT TCTAGACTGT TTACACACCT ACCACTCCTA
70741 AATGCACACA TATAATTGCT TGCACACACA CACATACACA CTCATCTCTT CTCTGGTGGT
70801 CCAGCTCTAT CTCTTATCAT TAGGCTTCTT GGGGCTAGTA CCTAGGGCCT GTATCCTTTC
70861 AGAGGCAGCT AAGGGAAGCA CACATAATTA GAAAGAATGA ACCAGCTTGT TGGATTTGGT
70921 CTCTTCGCAT CCAGCCCTCC AAGTTAAGGA GAGTACCATC TTTCTTAGGG TCACCAAAGG
70981 AAAAAAAGG AAAAGAAAGA AACAGAAGGA TATCATACAG CAAGGATCTA ATGCAAATAT
71041 GCCTCAAATG AGAGGCTACT GTGTGCTGAT CCCAATCCCA GGAAGTGTAT GCACATTATC
71101 TAATTTAATC CTCAGTGTAT TTCTGGGAGT ATTATTTCCA TTTTACAGAG AAGGAACTTG
71161 GCAGGGTAAC CAAGCTCATG AATGGAGAAA CTGGGATTAA ATATAAAGCT TCCTTGCTCC
71221 AGAACTGCTG TCTTTCTGCT CTTCCACACT ACCAGCTCAG CTGTGCTCTC TACATGCAGG

Figure 9 (Page 22 of 74)

SUBSTITUTE SHEET (RULE 26)

111/162

71281 CAGTTTTACA AGTTTCAGAT TAGCCTGGGA CTTCCAGGGT TTTGAATGGG TTAGGGGAATG
71341 GGGAACTTTT GGGTTTACTT TCCATTTTTT CTTACATACAT ATGTAATATA TAACATAAAT
71401 CTATGGTATA TATGATAAAT ATATGGCTAC ATATGAACTA TATAATCACA TATATGCATT
71461 ATAAATAAAT ATTAATTTTA TAATATTTTA AAGGTTATCA AATAAATATT AATATAAATA
71521 ATTAATAAAT TAATACTCAG CTTTGTTTTC CAAAGTGATA AATGCCTATA TTTAGCAAAA
71581 TATTTTTTGG AGGCCTGATA GTTTTTAGGA GTGTAAAGAA GTCCTGATAT CTAAATGTTT
71641 AAGAACCACT ATTTTAGGCT GTTGTCTTCT GTCTTATTTT CCCAGCTAGA CTGGTAAATA
71701 CTTGAAGGCA AACGTTTAGC CAGCACATTA ACATTTTATG TTTTATTTCT TTTGTGCTCT
71761 CAGTGGCTGT GTCTTTTCTA TCGATTTCTC ACACCTGTATG ATGGTTATAT TTGTCTGTAT
71821 CTGTCCCACC AGGTATAAGT TCTTGAGAGG ACACACTGCT AGGCTGATCT TAGTTTTTAT
71881 TATTTCTCCT GGTGTCCTGT GCTTAACAAG TGCTCATTAA GTGTGTAATA ACACAGCACA
71941 GTAAAAAACT AGACATTAAA AAATAATGTC AACCATCTA TTGAAATTG CATTTCATG
72001 TTTCTTCCAA TATAGTCATT GTGTCAGGTT ATGTACTTAT TCTGATGAAG ACTATTGCCT
72061 AATATACGTT TGCATCTTGT GCTTTATAAC TGCCTTCATA TAGACACAGA TTGAGAAGGT
72121 GTAAAAATGT GCATATCCTC ACAATTGACA AATTCTTATC CTTTGAGGGT AGGTTTGACT
72181 TTCTGAAATG CTTTGACATC ATTTGAAAGA AGCTTGAAGA ATAAGATAGC TGTTAATGAC
72241 CCAGTTTCCT ATGTCACTTA TACAATTATA ATGGCAATTT CAAAATGTTA GGTAAATATA
72301 TTTTGCAATA TATTGTTCTT TTTGTAATAC TCTCTATGTA TTTATTTATA TTTTAAATT
72361 TTATATTTAT GTATTTTATT TTCTGGACAG AGTCTTGCTC TGTTGCCAG GTTAGAGTGA
72421 AGTGTGTGTA TCATAGCTCT CTGCAACTTC AAACCTGCTG GCAAAAGTGA TCCTCCTGCC
72481 TCAGCCTCAT GAGTAGAGTA GCGGGAACCT CAGGCGCATG CCACTGCACC CAGCTAATCA
72541 CTATTTATTA TGCTCCTACT GTGTGCTTTA GTATATTTTC TGTTGTTTTC TGCAACCCAT
72601 TTTGAGGGCG TGTTAGGGAA TACAGATGCA GTAACTTTGG TCTCAGCCCT TGAGGTGAGG
72661 AAATATTTAG CCTCAGGTTT AATCTAATTG TTGGCCATTT GCCTTCAAAG ATTGAAATAT
72721 GAGCAAAACT GTGGCTCTGG GTTATATGTT AAAAAAAGT TTATGGGGCT GAAGCCAGGC
72781 AACAGACAAG AGCCCTTACA ATCTTATTTA GGCTGAAAAT ATCTGGAGT CCCTGTATTG
72841 TTGGTCTCAA GCAGATAGCA AACTAACAC TTAATCTTTG AGGCAGGCAC TGCCAGTGGG
72901 GTGGCTGTTA TTATTAGCTT CATTAAATGG TGAGTCAGGA AAAAACAGCT TTAATCATT
72961 CAAAGTTCTG GCCTATACAG GATTTAGTAA TATTAGGTTA GCTACATCCA AAAGATGACA
73021 GAACCCTACT CTAAGGCTGG GCTTGGTGGT TCACACCTAT AATCTCAAAA CTTTGGGAGG
73081 CTGAGGCAGG AGGATCACTT GGTGCCAAGA GTTTGAGACC AGCCTGAGCA ACATAGTGAG
73141 ACCCCTGTCT CTATCAAAAA CAAAGAACTC TAATTGGCAT AGTAGAAGGA AAAAGTGAAA
73201 GAAAAACCAG CTGTCACCCT CATTCCCTAC ACCTGTCTTA ACAACTCCTC TCACTATCCT
73261 TTGAATATAT CTTGGCTGTT TGAGTCTCTC TCTAGCCCCA TTAATGCTGT TTGGACTTGA
73321 CATTTTGCTC TGCAATTTTA ACTTTTCTAC CAGGGTTTCC AGACCTGAA GAGTGTGGCA
73381 TGAACAAAAA CTAGTCAACC TATAATATT ATGATGTGTG TGTAATAAAA AGAATACACA
73441 ATATATTGCA TTACAATATT TTAATGTGT CCAATTTG TTTGTGGCTT TCTTGAGGAC
73501 ATCAGTTTTG GGTGGGACGA CCACATCCTT AATCTGAACT TTCCCTTGA GGTCACTCTT
73561 TTTTTTTTGA AATAGAGTCT CGCTCTGTCA CCCAGGCTGG AGTGCAGTGG CGCAATCTCA
73621 GCTCACTGCA ACGTCCGCCT CCTGGGTTCA AGTGATTCTC CTGCCTCAGC CTCCAAGTA
73681 GCTGGGATTA CAGATGCACG CCACCATGCC GAGCTAATTT TTGTATTTTT AGAAGAGACG
73741 GAATTTCAAC ATGTTGGTCA GGCTGGTCTT AAACCTCTGA CCTCATGATC TGCCACCTC
73801 AGCCTCCTAA AGTGCTGGGA TTACAGGCGT GAGCCACCCC GCCCAGGAGG AGGTCTTCT
73861 AATAGACTTT TTTTTGTTG TTGCTCACAG GCTTGTTCAT TCTTATTTCA AAATTTGAGA
73921 AATACAGTTT CCATGGAACA CCAACCAGAT ATCAGGTTGC TATGGAGTTG ATAGTCAAAA
73981 GCTTTGTATC TTCCAGTTT TCAGAATGGC TTCTAAAGGT TCTGATTAG AGCTCTTAGG
74041 CGAAATTGAA CAACCAAGTG TCAAAGTACA ACATTCAGGA AGTTAAAAAC ATGACTGACA
74101 TATATGTACT ATATATAGTG AGCTTGTGTA TGTGTCAATG AATGATTAA TTCATTAAATG
74161 AAGGAGGAAG CAGAATCACA ATTAGGTCAA AGGAAGATAC GGGAGAATAA AATATGTATT
74221 TGGTCAGGGA AAGGATGTAT ACTGGAAGAG GAAGGGAAAA TCAGATATAA AGTTGTTTAA
74281 TGACTTATTA GGCAATACAA TAATAACTTT TAGGGTCATT TTTCTATAT TAAGAATTCA
74341 TTTCCATCTC TATGACAAAA TCCTTATTAA TTTATTAAAC TTCTACAAGT GAATGTTTAC
74401 TTTTAGATAG TCTGGACCCA ATAAATGTA AACATTAAGT CAGAGTTACT TTCACGTAGG
74461 ACAGTGTGTT CCAATAAGGT ACCACTAGCT ACACGTGATC ATTGACCATT TGGACTATAG

Figure 9 (Page 23 of 74)

SUBSTITUTE SHEET (RULE 26)

112/162

```

74521 CTAGACTGAT TTAAAATGTT CTAAGAAGTGT AAAATACACA CCAGGTTCTG AAGATTTATC
74581 ATTTAAAAAA GAATGTCAAC TGTCTTTTTT TTTAGCTTAT TTATTATATG TTGAAGTGTAT
74641 AATAGTTTAG ATATATTAAG TTAAATAAAA TATCTTAAAA TTAATTTTAC TTGTTTCTTT
74701 TCATTCTTTC AATGTGACCA CTAGAAATCT GGAAAGTATT TATGTGATTC ACATTCTATT
74761 TTACTGTCTA GTATTGCCTT ACATCATCAG GTACCCCAT AAGTAGGCTTT TTAGATAATT
74821 CTCTAATATA GCTTGGAAGG ATATGGAGAA ATATTTTTGC GTTGCTTTTA AGTTTTGCAT
74881 AACTTTTTCA ACACACTTTA TAAAGGATCT AGAAAAGGGT TGGTTACATG TTTCTCTGTC
74941 TTCTGGCCTC CACCATGTTG CCAGGAGGTT GGGGACAAGA TTCTGGGTGG CTGGATGTCC
75001 TAATGGCTTG AGGTCTGGAC TTGAGATTG CATATAAAGA GATGTGATTA GATTGAGTCG
75061 ACTAGAAAAA TCATATTAGA GAACGTAAAT ACAGCGATTA AATTTACATG TCGATTTATA
75121 AACCAGGACA CCAATTTATA GTGAAAGAAG GTCCAGTTAC CTGGTAATCA AGACGTTTCA
75181 TAGCTATTTT CATGATGGAT ATACTTAGCT GAGTTTTAAA TGAGAAGGGG GTTCATTGCA
75241 CATAGAATAA GATCTAAGTG AAATGTTTAT TTATTTTTTT TTTTTTTTGA CATGGAGTCT
75301 TGCTCTGTTG CCCAGGCTGG AGTGCAATGA GGCAATCTCG GCTTCTGGAG TGCAATGAGG
75361 CAATCTCGGC TTCTGGAGTG CAACGAGGCA ATCTCGGCTC ACTGCAACCT CCACCTCCCG
75421 GGTTCAAATG ATTCTCCTGC CTCAGTTTCC TGAGTAGCTG GGATTAGAGT TGCCTGCCAC
75481 CACGCCAGGC TAATTTTTGT ATTTTTTTTA GTAGAGATGG GGTTTCACCA TGCTGGCCAG
75541 GCTGGTCTCG AACTCCTGAC CTCAGGCGAT CTGCCCCGCT CAGCCTCCCA AAGTGCTAGG
75601 ATTACAGCG TGAGCCACCA AGCCTCGCCT AAGTGACATG TTCTTATATT GTTCTTTTCT
75661 TTCTTTTTTT TTCGACTGAG TCTCACCTCG TTGCACAGGC TGGAGTGCGA TGGCGCTATT
75721 TCGGCTCATT GCAACCTCTG CTTCCCGGGT TCAAGCGATT CCCTTGCCTC AGCCTCCTGA
75781 GTGCCACCAC CCCCAGCTAA TTTTGTACT TTTAGTAGAG ATGGTGTTC ACCATGTCCG
75841 CTAGGCTGAT CTCAAATCC TGGCCTCAGG TGATCCGCCC CCGAGTCTCC CAAAGTGCTA
75901 GGATTACAGG CGTGGGCCAC GGGGCCAGC CTTATATTAT TTCTTTTACT ACAATATATT
75961 AGTATGATGC AGGTGCTTCA ATTGTTTATA CACTTTCCAT AATTTTGTAT AATTCTTATA
76021 CCCTGTCACT CTGAGGAATA GCCGGTCTAA GTGTTTTTCC ACCACTGCTA ATTCATCCAT
76081 CACTAATCTC ATTAGACTGT TAATTTCCAG AGGACATAAG CACACAAGCA GACAATGTTT
76141 ACAAATGTTG GACAAATGTT ATTTAATAAA ACAATGGGGT CACCCTTAGT CTAAAAGATG
76201 TTTCACTTTT CATTTGTCTA TGAACCTTTA TTTGTAGGTT CCCTTTTGAC TTTCCACAA
76261 TCTAAGGCTG TTCTCTTAA CACATATTTT CATGAAAACA TATATTGAG CAGAAATTGT
76321 TGGGGAGTTG TAATATTACC TTTGTCCCTA AATATGAATC TATAATTATA TCAAATATAT
76381 GGGCAGACAA TTTACTTTGC CTTTAATCTC AAGAAAAAAA TAGCAATTAC TTGGGGTCCG
76441 AGAGTAAAAT AAGAAGTAGT GAACCTTAAA GTAGCAAAC TTAGAACAGA ATAGTTTCAG
76501 AGGGGATGAG AAGAGGTGAT TTTTCAGCTC ATCAACAACA GATCTTATAA TAAATTACAT
76561 GTTCTGGTAC TTTTCTTGTC TTTCTGTGTT AAATTTTGCT ATTTAAAAAA ATAAATTTCA
76621 AATACATTGT TCATCTTAA AGTCAAGAGT GTGTTTTATT AAAGTCAGTT GCTTTATTTG
76681 CAACTCAAAA GATATATTG AGTTCCCAAC TGGAGATTGT CCTATATGGT AACTTGCCTA
76741 AGGTATGGTT ACTGAAAGTA ACCTACAATT TTCATGGGCT GAAATTCATT TCTATATTGC
76801 AGCGTACAAA AATAAATAAA TAAAAAATGC TTGTTTTCTT TGAAAAACATA TTATCTCAGT
76861 GCCTCTAACT GCCAAATCTA TTGGCTTTTT TGCAGGCTTA AGGGCTCTCC CTTGTTCTCT
76921 TATGATCTCT ATCTTGAGGG CCAGACCTCC TGCCTTACAC AACTCAGAGG GGGACCTCAG
76981 AGCTCTTTAA AAAGAGCCCA ATTTCTCGCC TGTAGAGAAG TGAAAAGGAT GCCCCACCCC
77041 CATCTATGAA AAGAGGGATT TGATAGTTTC AATGCTTTCA AATCAAAGAT TTAAGTCTGT
77101 AGCCCCCACC CACCCCGGAC CCTAGCAAGG CTCATGAACC CCCTCCCATC CCGCCCTAAT
77161 TGCTTTGGAC TGGCCGTGGA ATCCTTGCTC CAGTCCACAG TTCCTGTGCG ACTGCACGAA
77221 GAATTCACAG AGGACCTGTG TTAATTCCCT TGTGAAGAAA CAGAATTATC ATGAAAATTT
77281 AGGTGGAAC CATTTGCTTT TTTTCTTCAA AAATAAGGGA AGCATGTGCC CAACCACCCC
77341 TGGGAAAAAG AACCTTCAGG GGCAAAGGAG CGAACAGGTA ATTTATAAGA AAAACAGAAA
77401 GTGTCTCTG ACTGCCCCAG ACTTCTTTCG GAGTTGGGGG AATTGGGGAC GCCTGGACGC
77461 GTTGTTTTTG CGTTTGTTGA AAAAATAAAT GAAGAGCATG AAGCCCGAGG CTTCTGAGAT
77521 CCTTTCCTGA CCAAACCCAA GTGATTGGT GCGGGGAATT TTAATATTTT TCCCCTTTTG
77581 TGAGGTGGAA CAAACACAAC TTGGGAGCAG CGCAGCGGCT CAGAGCCTGC CAGCCAGGCG
77641 GGCGACCAGA GCACCAATCA GAGCGCGCCT GCGCTCTATA TATACAGCGG CCCTGCCACG
77701 ACCTGCTTC ATCGCGCTT TGCCACTTGT ACCCGAGTTT TTGATTCTCA ACATGTCCGA

```

Figure 9 (Page 24 of 74)

113/162

77761 GACTGCTCCT GCCGCTCCCG CTGCCGCGCC TCCTGCGGAG AAGGCCCTG TAAAGAAGAA
77821 GGGCGGCAAA AAGGCTGGGG GTACGCCTCG TAAGGCGTCC GGTCCCCCGG TGTCAGAGCT
77881 CATCACCAG GCTGTGGCCG CCTCTAAAGA GCGTAGCGGA GTTCTCTGG CTGCTCTGAA
77941 AAAAGCGTTG GCTGCCGCCG GCTATGATGT GGAGAAAAAC AACAGCCGTA TCAAACCTGG
78001 TCTCAAGAGC CTGGTGAGCA AGGGCACTCT GGTGCAAACG AAAGGCACCG GTGCTTCTGG
78061 CTCCTTTAAA CTCAACAAGA AGGCAGCCTC CGGGGAAGCC AAGCCCAAGG TTA AAAAGGC
78121 GGGCGGAACC AAACCTAAGA AGCCAGTTGG GGCAGCCAAG AAGCCCAAGA AGGCGGCTGG
78181 CGGCGCAACT CCGAAGAAGA GCGCTAAGAA AACACCGAAG AAAGCGAAGA AGCCGGCCGC
78241 GGCCACTGTA ACCAAGAAAG TGGCTAAGAG CCCAAAGAAG GCCAAGGTTG CGAAGCCCAA
78301 GAAAGCTGCC AAAAGTGCTG CTAAGGCTGT GAAGCCGAAG GCCGCTAAGC CCAAGGTTGT
78361 CAAGCCTAAG AAGGCGGCGC CCAAGAAGAA ATAGGCGAAG GCCTACTTCT AAAACCCAAA
78421 AGGCTCTTTT CAGAGCCACC ACTGATCTCA ATAAAAGAGC TGGATAATTT CTTTACTATC
78481 TGCCTTTTCT TGTCTGCCC GGTACTTAA GGTAGTCTG ATGGGAGTTA CTGAGGTATC
78541 AGAGACGAAT TGGGTGACGG GGTGGAGAG TGGCCGTGGT GAGGTTACAG CATTTAAACC
78601 TTTATTGCGG CTTCTAGGTC CCTGACCGGA GGCTTTTCTC GCTGGCGGAT GGTTTGGGA
78661 TGGCAGTCCC GCCCAGGCC TGTGAACGGC AGAAAAGACC GCAAAACAAG AGCCAGTTTC
78721 TTAGTCTAAA GGGATGTCCG GATTGGAATA AAAAATTTTC AAAAGTCCCG CCTGCTCCC
78781 GGGTTGGTCC GTTCTTCTAG TACATGACTT TCATTCTGTA TTTAATTGGA TGGTGAAGA
78841 CGTTGCTTAT TCTGTGTTTT TTGCTTTACT GTGACTTAAA AGTTTTGCCT CTTTTCTCTT
78901 TATATTAATG TCTGGGATTT CGGACGCTTT CCATGTTGTT GGTAGTCAAG TTGATGTCTC
78961 CTGGAGGTAG TGGCAACATC CAGCCCTGGG AGGAGAGTGC GTGCAGGTAC CTTTGTCTTA
79021 CATTCCTCTG CTGTTAATTT CTCATCTCTG TGGCAACGAA GGAATGCATT TAAAAACAG
79081 CCACAACAGC GGCAATAGCC CTTCTCCAC CCAAGGCAAT CGTGGACCTA GGGAGTTTTT
79141 TGTGCCACAT AACATGTAGC CTTCCGCTAA ACTGACAGGT TTGAGCGTAT CGATTTTGAG
79201 CGTATCGAAA GCACAACCTT TAGCCAGCCA TTTTGTCTC GCATGACTAC GGTGCTTAT
79261 CCTGTTTAGA CAGACAGCAA CATTTAAAA TCGAAGTTCC TTTAAACGTA TTTTGTTTGG
79321 CAGTCCAAAT GTTTCTATGC AGAAAACAGT ATTTGTACTA TTAACATGA AGAGTGTATG
79381 GATAAATGGG AGACATTTCT AATAAAGGCC TTCGTTAATG GTTCCCTCTG TTTGACATCC
79441 ATGGTGCTTC TGAATACAGA AAGCCTAGCG TCTTATATTC GCTTCTTTTA AAATCTGGTG
79501 GGCACATTTT GGTGAGACCT AAATTATGGG GACTGGGGCT TCTGGAGATA AGCTGCTCAA
79561 TTATTCTACC ATCTCCACAA TGATTAATAT AGTGAGTTGA TTTGTTAGTG ATAGTGACCA
79621 CGGATTCATC CCAAGAAAGA GAAAGGGGAG GGAGGCAAGC AGAGAGACAG GAAGACAGAG
79681 GCAGGGAAGA AGGAGAAAAC ATTCTCCCAT GGTTTAAGTA ATTTGTGTT GTTAATTTTA
79741 CATTACAACA CGGTTTAAACA TGGTGAACCC TCTATTTTGG TGAAGGTTT AACATATGGA
79801 CATATTTTTT CCAAGACCAT TTATGAACTT TCATTTCTGC TCCCCCTTC TTCCTCCCGT
79861 GCCACCCTCC ACGCTCCTAT CAATTTTGGC TGTTTTGTC TAGGCTAATA CGCTATAATT
79921 TCATGGACAG TTGGACTGTC TTAGGTTTCT CAGGTTTCTA TTTGTTCCCT TTAGTCATTC
79981 CCACAATTCT TAAGGTAGAA TTGTATTGTT TTAACATTG TGTGTGTGTC TATCTCAAT
80041 GCTGAGATGA TTATGTGACA AATGGCAAGT GTTCAACTAA TACCTAAATC TGATGATCT
80101 TATCAAGCCT AATGCTACTT CACAATGCCT ACTCCATTCA CCGCACTTTA TCTCATTACT
80161 GGCATTCTGT CATCTCACAT CATCACAAGT AAAACGGTAA GCTATTTTGA GAGAGATCAC
80221 AGTCATATAA TTATATTTAT ATTTATTTAT TTATTTATGA GACGGAGTTT CCCTCTGTCA
80281 CCCAGGCTGG AGTGCTGTGG CACGTTCTCG GCTCACTGCA ACCTCCGCCT CACGGGTTCA
80341 AGCGATTCTC CTGCCTCCGC CTCCCAGTA GCTGAGATTA CAGGGGCTG CCACCATGCC
80401 CGGCTAATTT TTGTATTTT AGTAGAGACG GGGTTTCACT AAGTTGGCCA GGCTGGTCTC
80461 GAACTCCTGA CCTCAGGTTA TCCGCCCACC TCATCCTGCC AAAGTGCTTA GATTACAGGC
80521 GTGAACCACC GTTCACAGAC TCAAATCATT TTTATTACAG TATATTGTTA TAATTGTTGT
80581 TTTATTATCA GTTATTGCTA ATCTCTTACA GTGCCTGATT TATAAATTAA ATTCATCATT
80641 GCCATGTGTA TATAGAAAAA AACAGTGAT ATACGGTTCA GTACTATCTG TGGTTTCAGG
80701 CATCCACTGG GGGTGCAATT TATTAACAT GCATTTACAT TAGTCTCCCC TTTGGGAGAC
80761 TAATTAACAG AGATGTTGTA ACGTGACTTT AATAGCAGAT AGAGCTAATT TTCTCTCATT
80821 ACTCTTCTTT TTCAGAAATTT TCCTGGTTAT TCCATTTTTT ATTTTCCAT ATGTATATTA
80881 AGATCTCTTC CACCTCCTCC TGTTTCTCCA TCTCAACATC AAACAATTAA AAAAAAATA
80941 AAAGCTGGG CGCGGTGGCT CACGCCTATA ATCCAGCTC TTTGGGAGGC CTAGGCGGGT

Figure 9 (Page 25 of 74)

SUBSTITUTE SHEET (RULE 26)

114/162

```

81001 GGATCACGAG GTCAGGAGTT CAAGACCAGC CTCGCCAAGA TGGTGAAATC CCGTCTCTAC
81061 TAAAAGTATA AAAATTAGCC AACCATGGTG GCAGGCGCCT GTAATCCCGG CTAAGTCGGGA
81121 GGCTGAGGCA GAGAATTGCT TGAACCCGGG AGGCGGAGGT TGCAGTGAGG CGAGACCTTG
81181 CACTCCAGCC TGGGTGACAC AGCGAGACTC CGTCATAAAA AAAAAAGCCG GAAGCAGTGG
81241 CTCACGCTG TAATTCCAGC ACTTTGGGAG GCTGAGTCAG GCAGATTACC TGAGGTCAGG
81301 AGTTCAGGAC CAGCCTGGCC ATGAAAATAC AGCCTGGCCA TGAAAACACA CAATAAATTA
81361 GCTGGGCGTG GTGTCACACA CCTGTAATCC TAGCTACTCG GGAGGCTGAG ACAGGAGAAT
81421 CACTTGAACC CAGGAGGCAG AGGTTGCAGT GAGTTAAGAT GACGCCACTG CACTCCATCT
81481 GGGCGACAGA GCCAGACTCT CTCTCAAAAA ACTAAATAAA TAAAAATAAA GTTATGGTAC
81541 ATTGAACCTC TGTGTTCCCT TCTCCCTTAG ATACTTTCAT GGCTACCCAT TTAATTGATG
81601 TTCTTATCAT CTCCAAGAGT TAGTCAGGAG AGGAATCAAC CCAAGCAAAA ATAGGTGATT
81661 TTCTAATTTT CCTTCAATGC CCTTTGGGGT CTTAATCCAT TTGATTTATG TACTTTCAAT
81721 TAATCCTAAC CTCGAATGTC TTCTGCAAAC ATGTTTCCAC AGATGAAACT CGTCAAATGA
81781 AACACATTCC TTTAATTTAT AGAGTTAAAA ATTAGAAAAA TTTTCAATTC TATTTGGCCT
81841 TTAGATTGAG TCTTGACATAT GTTTTCTCAA TTTTGTTCAT GCTCTTTAGT TTTGTTTTAT
81901 TCCATCACAA TTGTTTACAT AGCTTACTGG CTTAGGTCTA ATGAACCATT CATTTGGAAA
81961 TTAATAATTG CCATTTTAAAG ATGAAAAGA TTCTTGCTC AATTTTACTT AGTTTTTGAA
82021 ACTGTCAATG AGGACACATG TTTTCTGTA CTCTTAGATT CACTAAGTAG TGTCTTGCAA
82081 ATTTAACTGA CAAAGGACAG ATTAACATGC GAAAAAATAA GCATGCAATT TTATTAGTAT
82141 ATTACATGCA CAGAGTTCCC AAAGAAAAAA AAATTGAAAC CTTAAAAACG CGGTAGACT
82201 CACAGACTTA TACACCATTC CAACAAAGGA AAGGGAGTTT GCACCTTCATG GGATGACGAA
82261 TTTGGGAATG TGACAAGGAA ATAAATACAT GGGCAATAAA AACCATGGAA GATAAATGA
82321 AAGATAGAAA TAATTGTAGT AAGGTTTGT TTTGCAGAGT CATCTCAGTG CCAACCTTCC
82381 ATATCTAGTG ATAAGAATTG CTCTCTTTT CCTGGTATAG CAGTTGGGGA CACTTTTACA
82441 AGGGAAATTT CTGTCACCTT CACAAAGGGA AATTGGGTA AAGAGAAGAC AGAGACCTCT
82501 TCCTACACCT GTTGATTTTC AATTGCCTTC AGCTGAAAAT AACTTTTATG CCAAAGTAGA
82561 ATAATTTGGG GGTGACATCC TGATATTCTT CAAAACCTAT ATTTAATTTT ACATTAGTAA
82621 TTATATCATT TTTGATTTTT AAATTAGTTT TATAAATAA TTTTGAAAAA CGGTAATAAT
82681 ATTCAAATAA TTCCAGAAAC ACTGCTGATA AGCCAAAAAC ATCAATGAAT ATTGCATAAA
82741 CAACTGATAA TTCAACCATG AAAATTTATG ACATTGTTCT TGTGTGATAA AACTATGAGT
82801 AACATAAAAA CTAGAGGCTA CTGTAATGC ATTATTCCAA ACTTCTGTT TTTTATTTAT
82861 TTATTTATTT ATTTTGAGAC ATAGTCTCTC TCTGTCACCC AGGTTGGAGT GCAATGGCGT
82921 GATCTTGGTT CACTGCAGCC TCCACTTCCC CGGTTCAAGC AATTCTCCTG CCTCAGCCTC
82981 CTGAGTAACT GGGATTACAG GCACCTGACA CCAAACCCGG CTAATTTTTT TGTATTTTTA
83041 GTAGAGACGG GGTTCGCCA TGTTCGCCAG GCTAGTCTCG AACTCCTGAC CTCAGTGATC
83101 CACCTACCTC GGCCTCCCAA AGTGCTAGGA TTACAGGCGT GAGCCACCAT GCCCGGCGCA
83161 TTATTCCAAA CTTTCATACA CAGTGCTATC ATGGCTACAA ATTGAAGTAT CATATTATAC
83221 ACTCCTAGGC AAAGCTCTGG ATATTTTGGC TATATAAGCC TGAGGGAAT GTAGTAAGGA
83281 CATTGTGGTT GAAATTCATA CCAGAGATGA ACAGGCCAG TGCAAGACAG AATTACATCA
83341 CTAAAGGATA TCAGAAGAGA ATAGGGATT AGGGTACAGT GGCAACAACA GTTTTGGGAA
83401 CTAGCATTTT TTGAGCACTT ATTTACAATA TGCCAAGCAC TGTGTCTGAT TACTCTATAT
83461 TTATTTTCAA ACACATTCTT GTCACAGCAC TTTGAAGTAA GTGCCATTGT CATTCCCACT
83521 TCAGGGTGAA GGAATAAAGC TTGGTGTGAT TAAGGATGTA GCTAGTTAGC TGTGTGTGTG
83581 TGTGTGTGTG TGTGTGCATT TTTTTTAAA TTTAAAGTCA ATAAATTTT ATTTGAAGAA
83641 TTTACATCA AGGTAACTT TGTTCCTCTA AAGAGCTGGA GTCAAAATGT ATCTTCAAAA
83701 GATTATCTT CAAGTTAGCC CTTCTTAATA GAACTGATGC TTAATCCACA GTTGTGAGCC
83761 CACAGTTCTT TTATTTTGAC TTTTTTTTTT TTTTTTTTTG AGACGGAGTC TCTCACTGTC
83821 ACCCAGGCTG CTGGGCGAGT GCGTGATCTC GGCTCGCTGC AACCTCTGCC TCCCGGGTTC
83881 AAGTGATTCT CCTGCCTCAG CCTCCTTAGT AGCTGGGACC ACAGGCGCAT GCCATCGTGC
83941 TCGGCTAATT TTTGTATTTT TATTAGAGAC AGGGTTTCAC TATGTTGGCC AGGCTGATCT
84001 CAACTCCTG ACCTCATGAT CCGCTGCTC TGGCCTCTCA AAGTGCTGGG ATTACAGGTG
84061 TGAGCCACTG CACCCGGCCT TATTTGCTT TCTTAATCT CCATTGAAAC ATGGACATAC
84121 TGATGAAAAC TACAACATTC TTCACAAAA ATCTTTGGGA TTTAATTTCT TCAACCACTT
84181 TACTTTGGGG TCATTTTAAG ATTAGGTGTA TCTGCCTGGT TCTCAATTTG ACACCTTTTC

```

Figure 9 (Page 26 of 74)

SUBSTITUTE SHEET (RULE 26)

115/162

84241 TCTCTAAACA TGAATGAGTT CCAATCATAT TTATTCCTAA GCTATCACAC TCAAATATAC
84301 TACAGATCTG TGGAATATGC CAAAAGTTAA GGTGAAAAAT TAAATTATTA GGTATTTTCAT
84361 AGTTTTGCTA GTTTTTGATC TGTGAGTGAA TATAACTATC CTCTATGTCC TGGCACTGTT
84421 CCTCAGAAAC ATAGGGTCCA CATATGTAAT TTAAATTTT TTAATAGGCA CATTTTAAAA
84481 AGTGGAAAAA GAAATCTATT TTAATGATT GAATCCAGTG TAACCAAAAA TTGTTTCAAC
84541 AAGGTATCTA ATATTAAAT ATTGAGTTTT TACTTTGTTA TTTTACTAGG TCTTTGAAAT
84601 CTGGTGTGTA TTTTACACTT AAAGCACATC ACAGTTTGGG GTAGCCACAT TTCCAATGCT
84661 TAATACTCAC ATATGGTTAG TGGCAACTAT CTTGGACAGG ACAGCTTTTA TACTCTGGGA
84721 AGACACAAGC AAATACTTGC TCTGCAGCAG AATCCAGATG TTTTCCAAGA AAACACTTTT
84781 TCTGACCTGT TCGTGAAACC CAGGTAGTGT CTCTAATACT TTATATTTTA TTGGTTTGTG
84841 CTATTGTAAC CACCCAACGG GCTCTCCTTG TCCACTTCCT AGACAGAGCT GATTTATCAA
84901 GACAGGGGAA TTGCAATAAG GAGCCAGCGC TACAGGAGAC TAGAGTTTTA TTATTACTCA
84961 AATCAGTCTC CTTGAGAATT TGGGGACCAA AGTTTTTAAG GATAATTTGA TTGTAGGGGA
85021 CCAGTGAGTC GGGAGTGCTG CTTGGTTGGG TCAGAGATGA AATTATAGGG AGCCTAAGCT
85081 GTCCTCTTGT GCTAAATCAG TTCCTGGGAG TGGTGGGGTG GGGGACTCAA GACCAGATAA
85141 TCCAGTTTAT CTATATGGGT GGTGCCAGCT AATCCATTGT GTTCAGGGTC TGCAAAATAG
85201 CTCAAGCATT GATCTTAGGT TTTAAATAG TGATTTTATC CCCAGGAGCA ATTTGAGGTT
85261 TAGAATCTTG TAGCTTCCAG CTGCATGACT CCTAAACCAT AATTTATAAT CTTGTGGCTA
85321 ATTTGTTAGT CCTGCAAAAG CAGTCTGGTC CCCAGGCAGG AAAGGGGTTT GTTTCTGAAA
85381 GGGCTGTTAT TGTTTTTGT TAAAAGCAAA AGTATAAACT AAGCTCCTCC CAAAGTTAGT
85441 TAATCCCAA CTCAGGAATG AAAAGGACAG CTTGGAGGTT AGACGTTAGA TGGAGTCGGT
85501 TAGGTAAGAT CTCTTTCAC GTAATAATT TCTCAGTTAT GATTTTTGCA AAGGCAGTTT
85561 CACTGTCCAC TTCACCTCAC ATCAGGCCTC TGACTAGAGG ATTCCAACA TACTTAGGCC
85621 AGGACACCAC CATGTCTCCT TATCCACCCT GAGGGATTCC AATTTCTGAA ACAAGGAAA
85681 CTATATATGA TAGTATGAAA CTATATATGA GAAGGAAATT ATATATGATA ATCAATTTTA
85741 GGGTTATCTT ATTGATTAGA AGATATTTAA GTGTGACACT GCCTGGCAAT GATATCTGCT
85801 GGTAGTAAGA ATTTGGCGAA TTTAGTGAAA TTCCTGAGGC TGAACCTCCA CTCTGTAAAA
85861 ATGGAGACAG TGAGATAATT TGCTTACAA TGCTGAAGTA AGAATTTTAC ACAATAATTC
85921 AGACCAACCA CTTTCATGTG TACTTGGCCC GTGGAAGACT ATCAATGACA GTTAGTTTAT
85981 AGTTTTACT ATTAATGAAT CCTTGTGTTT ATTGTTATTT CCTTCTACAC GTTGGCCTCT
86041 CTAAAAGAAG GTAATATTCA ATACAAATAA AGTTAAAACA GCTTGCAGAG TTGTCCAGG
86101 GAACTCACTT AACCACGTAA GTGTTCAAAT TGCTTAAGGT TGACTTTATA TTCTCCTGAC
86161 TAACCTTTCT CCTTCTGGTA TTTCTTCTGA GAACAGCACC ACCATCCAA GCATCATGCA
86221 AACAGTGGTC ATCCAGACC AGTAATTCTC AACTCACAGG GTGCTCCTGC AGAGATGTAT
86281 TTGAATAGAG TGGTAGGATG CTGAAGAAG CCACGTAAAA TTTGGCCAGT GATCTGGGGC
86341 AGATTTATCC TGAAGCTAAT GAAACACAAG TGTAAGGGCC TGTAATTCCA AGGTGCAGAG
86401 AGGGGCCCTA CAAATGTGTT AGTTTGTCTC TCTCTCTCTC TCTGATTTTA AAATTTGCAG
86461 TATTAAGGTA CTTAATCAC GGATGGTTCA GGCTGCTATT TTCACTCAAT CCTCCTTTT
86521 ATTAATAATCA CCATTGTCTG ATTATGTTAG AATCCTGATG AAAATATTTG GAATTTGAGT
86581 AAGAGAAAGT TTAGTTGAAG ATGATCTAG TATGGGGATA ATAAGTTACG TGATTTGCAT
86641 ATGTGATCAT GTGTACTTCA TTCGTTGCCA GCCAATCTGA CGTAAGAATG GCTTCAAGGA
86701 GGCCGGGCGC GGTGGCTCAC GCCTGTAATC CTAGCACTTT GGGAGGCCGA GACGGGCGGA
86761 TCACGAGGTC AGGAGATCGA GACCATCTTG GCTAACACGG TGAAACCCCG TTTCTACTAA
86821 AAATACAAAA AATTAGCCGG GCGTGTGGC GGGCGCCTGT AGTCCCAGCT ACTTGGGAGG
86881 CTGAGGCAGG AGAATGGCAT GAACCTGGGA GCGGAGCTT GCAGTGAGCC GAGATCGCGC
86941 CACTGCACTC CAACCTGGGA GACACAGCGA GACTCCGTCT CAAAAAAGAA AAAAAAGAA
87001 TGGCTTCAAG GAATGTTCTT ACTGCTCACT GGAATAACTC ACCTAAATTC CTGGCAAGAT
87061 GCAGGTCTAG ATAAATGTT ATGACATCTA AGTATTCAA ACACATTCCC AGCACTGAGA
87121 GTGAGTGTCT AGTGGAGAGT AGAAACGTAT AGAGCCAGAA GCTAGTCTGG AAAGAATTCT
87181 TACAAAGTTT ACAACTTACA TGTGAAAGGA GCTTAACAGA GGATTTTCCA AATTTGAAAA
87241 CAATCCTAAA AACTTACTTG ACATTACCAA TAATGTGTTT TGAAACTGAA ATACTTCTAA
87301 GTTATGAAGA AAACATATTA TCATCAGCCA CCCTGGAGGA AAGATTGAAT TCTATTTCCA
87361 TTACCTATAG ACAACATTAC AAAATAATTT CGATCTGAAG ATGGAATCAG AGTATTCAGT
87421 CAAAACCTACA GGAAATATA CTTGGTAGTG TCATATTCAG AAGTTAATAA AATATGCTAT

Figure 9 (Page 27 of 74)

SUBSTITUTE SHEET (RULE 26)

116/162

```

87481 TTTCTGAATT TTGTGATGGC TGTGTTTTG TCAGCTTTTA TAAAATTGGA ATTTGATTTT
87541 ATTTTCCCAT TATAAATTTA TATTTACAGT CTGCAGTACT TTTGCATTTT TAATTTTACA
87601 TTATAGCTTT TAATAGTTAA CAAGTTGTAA AAGGTTTGAT CCCAGAAAA CCTTGATCTA
87661 CCCCTCAGT TAAGTATACT AATATATTTA GAAAATGGAT GAAATCAGCA TTTGAATATT
87721 TTTAAATATT TATTAAAAAG GGACATGGGT AAAAGAGCTT TGCAGTTGCC ACCCTTCATT
87781 CTCAAATTCC CTGGATAAGG ATGACCGCAT AATCTTTGGA TGGTCATACG CAAGTCTTGT
87841 GTATTTGTGA CATAAATCTA TTTAGTGGAC TTTTGGCAGT GTGTACTGAG GCCAGTTTCT
87901 TCCACCTGAG CTCTGACTCC ACCTCCAGCA GCCCAAAACC AATACTGAAT TTTGGGGTCA
87961 GCTATTGTTT TTGTGGACTT AGGTAACCTAC ACACACATTG TCTTTATGAT AGCTTTAATA
88021 ATACTGCCAT CAGAACTAAA ATTGTCACGT GGATTAAAAG GAGTGACGGT GGTGTCCCCA
88081 GGAGCCTTTC AATATGTAAG TATTTACACA TATACATGCT AAAAAGACCC CTAGGAATTT
88141 TTTTAACAAG GGCAAAACAG TAACTCAGCT TGTTTTCTCG CAGTAAACC GGTGAAAAG
88201 GCCTGATAGA CTTGTCTGCA GTTACAAAAC TTGTGTGTAG TTATCACCTT TATATCTCCT
88261 GGAACTAAC ATAGACAACC GAATGGGTTA CAACTGTTT TAAGTGAAT TGTGAGTGGC
88321 TCTGAAAAGA GCCTTTTCAA TGAGGAAGAA ACGGGCAGAC TTATGCCCTT TCCCCACGGA
88381 TGCGACGTGC CAGCTGGATA TCTTTGGGCA TGATGGTGAC GCGTTTAGCG TGAATAGCGC
88441 ACAGATTGGT GTCTTCGAAG AGTCCCACCA GGTAGGCCTC GCAAGCCTCC TGCAGCGCCA
88501 TCACCGCAGA GCTCTGGAAA CGCAGGTCGG TTTTGAAGTC CTGGGCGATT TCTCGCACCA
88561 GCGCTGGAA CGGCAGCTTC CGGATCAGCA GCTCGGTGGA CTTCTGGTAG CGACGGATTT
88621 CGCGCAAGGC CACGGTGCCC GGGCGGTAGC GATGAGGTTT CTTACGCCA CCGGTGGCCG
88681 GAGCGCTCTT ACGGGCTGCT TTAGTAGCAA GCTGCTTGCG CGGAGCTTTG CCGCCGGTAG
88741 ACTTGCGAGC TGTTTGCTTC GTACGAGCCA TTTGCAATGA GAGCACACAC AAAAGTGTAG
88801 TGAAGTGAAG GCAAGTGGCC TTTAAATATA GTGAGAAACA TTCTGATTGG TCCTGTAATA
88861 TTTCAAAGT CCGCGCGAT AAAATCATTG GCTGAAGAGT GACCAGATG ATTGGTTTCT
88921 TACTAGACAA TCTTATTGGA TGAGTTGCC CACCGCCCAT CCTGTCCTTT TCGTTTCAGT
88981 TATCTGCAGC GACAAATTGT CTAAATTTCT AGTTCATCCA GTCCCAAAGA ACAGAGTGTA
89041 TAACAAGGTA TCTAAGGATT TTTAAATGT AAATCCGAT TCAGTAAGTT TGAGTGGGAC
89101 TTGAAATTCT GCATTCTCTGA CAGTCTCGCA AGTTATCAAT GCTGGTGAAC ACTCACTAAA
89161 CCACCAGAAA CGTTCAGACT CATGTCGGGA AATAACGCTT ATATTACAG AATGAGATTC
89221 CATGCTATTT TGTTACTGGC GAACAGCAAG TTTCTTGCC CTTTGTTTTC TAAGTCCAAG
89281 TCACATTTCC ACCCTGCCTG TTCTCAAAT GTCTTATTTT GGTGGCCCTT AAGTTTCACT
89341 TTGTATACTC TAAAATGTAC TTTCTAAAGG AAGGTGTTAT TTTCTCGAAA CTTAACTTTT
89401 TAACACCATT AGGCTAGGGG GGGCGTGCT CACGCTGTA ATCCCAGCAT TTTGGGAGGG
89461 CGAGATGGGA CGATCACTAG AGGCCAGGAG TTCAAGACAA CCCTGGCTAA AATGGTGAAA
89521 CCCCGTCTCG CATAAAAATA CAAAACTAG CTGGGCGCGG TAGCAGACGC CTGTAATCCC
89581 AAGTACACAG GAGGCTGTGG CATGAGAACC CGGTGAAGCG GCGGGGT3GA GGTGTCAGTA
89641 AGCCGATATC GCGCCGCTGC ACTCCAGCCT GGGTGACAGA GCTAGACTGT CTCAAAACAA
89701 ACCAATCCAA ACGAAAAGCA AAAAATACCC TAACAGAAGC AAGTTATCAT CCTTCTTGT
89761 GTAACATGAG ACGGCTCTGA AAAATGCCGT TTCAAGTGTA AGCTACGTTT TCTGATTTGA
89821 GTGTTTACTT GACCTTGCC TTATCGTGGC TCTGTTATTT TGGCAACAGG ACGGCCTGAA
89881 TATTGGACAG GACGCTCCC TGAGCAATAG TGACGTTGCC CAGCTGCTTG TTGACCTCCT
89941 CGTCGTTTTC GATGGCCAGC TGCAGGTGGC GGGGGATGAT GCTGCGGGTC TTGTCACGTA
90001 TGGCGCTGCC CACCAGTTCT AAGATCTCGG CGGCCAGGTA CTGTAAGTAC ACTGGCGCAC
90061 CGGCTCCGAC CGGCTCAAAA TAATTGCCCT TTCGAAAAAG ATGACGGACT CTGCCCTATT
90121 GGGAAGTGA AGCCCGGTAG CGACGAACAA GTTTTGTCTT TAGCTCCATT TTCCACGTCC
90181 GCAAATAGCG ACCTATGAAA GCAGCGGAAA ACTGTGAAAG ACAAGCAAGC TGGAATGGCG
90241 CCTGAACAAA TCCTTTTATA CAAACTGCAA GGCTGCAATA GGAAGCTATC CTATTGGTCA
90301 ATTATGTTTG GTGCTTTATC CAATAGAAAA AGATAACATA AATTCCATAT TTGCATAAAC
90361 CCCACCCCTC AGTGAAACCG TGTTTCTTTT GTCCAATCAG AAGTGAGGAA TCTTAAACCG
90421 TCATTTGAAT CTCAGGACTA TAAATACATG GGCTCTGAAC TGTTCTCTGT ACTACTCTGT
90481 AGTGGAGAGT GTTAGTAGCT TTTCTATTCT GTTTAGGAAT AGCAATGCCT GAACCTCTTA
90541 AGTCTGCTCC AGCCCTAAA AAGGGTTCTA AGAAGGCTAT CACTAAGGCG CAGAAGAAGG
90601 ATGGTAAGAA GCGTAAGCGC AGCCGCAAG AGAGCTATTC TATCTATGTG TACAAGGTTT
90661 TGAAGCAGGT CCACCCCGAC ACCGGCATCT CATCCAAGGC CATGGGGATC ATGAATTCCT

```

Figure 9 (Page 28 of 74)

SUBSTITUTE SHEET (RULE 26)

117/162

90721 TCGTCAACGA CATCTTCGAG CGCATCGCGG GCGAGGCTTC TCGCCTGGCT CACTACAATA
90781 AGCGCTCGAC CATCACCTCC AGGGAGATTC AGACGGCTGT GCGCCTGGCTG CTGCCTGGGG
90841 AGCTGGCTAA GCATGCTGTG TCCGAGGGCA CTAAGGCAGT TACCAAGTAC ACTAGCTCTA
90901 AATAAGTGCT TATGTAAGCA CTTCCAAACC CAAAGGCTCT TTTCAGAGCC ACCTACTTTG
90961 TCACAAGGAG AGCTATAACC ACAATTTCTT AAGGTGGTGC TGCTGCTATT CTGTTTCAGT
91021 TCTAGAGGAT CAACTGGAAT GTTAGCGAAG ACAAGTTTTA GAGCCAAGGT TAACTTGGAC
91081 GGGGCCGTGC GCGGTGCCTC TTGCCTTTAA TCCCGGCAAT TTGGGAGGCC GAGGCGGGCG
91141 GATCACTTGA GGTGCGGAGT TCGAGACTAG CCCGGCCAAC ATGGCGAAAG CCCGTCTCTA
91201 CTAAATACA AATGATAGAC GGTGCTGATG GCGCTCTTTC TCATCTGTCT TAGCAAACCTT
91261 CTTTGTTCCT CCTGGGTAAG CCTTCGGGTA CTATGTATAA TTCCTTTGAT AAGGTCACCTA
91321 CTCCCTCCCT GGTCTAGTAC AGGAACTTC CCTTCTGGA TAATGAAGCA GGTAAATGGAA
91381 TTCAGGGTAT AGTGTTCCTG TGGGGGTCAT TAGCCGTTAA CTTCTTGTA GATGCGGGGG
91441 AGGGGAGCAG AAAAGTCTAA GCGACAAAAG GGCATGTAGG GATATTTGCT CCTGCAGCTT
91501 GCCTATGCTG TAAATCTTA CTTCAAGTAT TGAGGAAACA ATAAGCGAAG TCTGATTTCC
91561 CGGGCGCCTT TATACGGAAT ATTTCCCGCT CCACAAAATG AAATCGCAGT AGTTTGTAGT
91621 TATAATTGTT TATCAATGAC AACAGCTATG TAGTTTACAT ATTTCTAGCA TCCCAGAAAT
91681 CCAGATTCCC ATTTCTTAAG CCACTTAACG TTCTGATTTT CAGCTCTGCG AGATACAAAA
91741 GGGTTTGGAT TTTGTGCCCT TCCCATCTG GCGCCACTGC AAAGCTTACT AGGAGGGCCC
91801 CACTTGGAGA GGGAAATCTT TTTTCGAGAAG TCCAGGACGC CAAAAACAAT ATAGCTAAAA
91861 AAAAAAAAAA AAAAAAGGCA GGAAGAGCAC TAGTTGAGGA GGAGGACTCA ATGGGCCAAT
91921 TCTGGGGCTG GGGCTGGGGG AAGAAATGCA AGAAGAAAAG ACCTTGTGTG ACTGCACAGT
91981 AAGCAGGAGG GGGTGGGGGA ATCGGAGGGG AGTATTTTCA GCGAATTTAT GGGCATTATA
92041 TGAGGTGAC ATACAGCAGT GTCTTTGGAT GAAGAAATAA AGTTTCTCAA ACAGTTCTTG
92101 TTTTGTGTTT GAGAAAGGGC CTTTCTCTGT CGGCCAGGCG CCATCATAGC TCACTGCAAC
92161 CTCGACTTCC CCAGCTCAAG CGATCCTCTT ACTTCAGCCC CTTGAGTGGC TGGGACTAGA
92221 GAAATGCACC ACCATACCCA GTTAATTTTT TAATTTTTTG TGGAGGCAAA GGGTCTTACT
92281 TTGTTGCCCA GGCTGGTCAA GCGAACTCCT GGGCTCAAAT GATCCTCCCG CCTTGGCCTC
92341 CCAAAGTCCT GGGATTATAG GAATGAGTCA CCGCGCCCGG CCCAGATTTA ATTTTAAAGA
92401 ATCTTTTAAA AGAGGTTCTG GGCCTGGTGT GGTGACAGTC ACGCCTGTAA TACCAGCATT
92461 TTGGGAGGCC AAGGTGGGAG GATCACTTGA GCGCAGGAGC TCAAGACCAG TCTGGGCAAC
92521 TTAGTGAGAC CTTTGTCTC CACCAAAAAT TAAAAAAT AACCAGGCCT GGTGGCACAT
92581 TTCTGTAGTC CCAAGTACTG GGGAGGCTGA AGTGGGAGGA TCATTTGAGC CTGGAAGGTG
92641 GAGGTTGACG TAAGCTGTGA CGGCACAAC GCACTCCAGT CTGGGTGAGG ACAGACCCTG
92701 TCTCAAAAAT AAAAAATAA AAAAAATCTG GATGCCACAC AAAATGTCTG TGAACAACATG
92761 TAAGTGAAGC ACTTCCCATC CTAGTACTGT ATATGCAAAC TGCCGTTGTG AAAGTGACGC
92821 TTGGCTTAAA AATCTACATT CTTTTTTTAA TTATAAACT ACCACATCCC CAAAAACAT
92881 TACTAAGGAA TTGAGGCTGC AGTTTAAGAA GCTGATATT AGGATCTATC TCCGAGAAAG
92941 TGAGACCTGG TAATATAAGC ATTTTCAAAA TGAATTTTG GGCCAGGTGA GGTGTGTCAT
93001 GCCTGTAATC CCAGCACTTT GGGAGACCTA GTCAGGCAGA TCACTTGAGC TCACAATTCTG
93061 AGACCAGCCT GAGCAACATG GCGAAATCCA GTCTCTACAA AAAATTAGCA GGGCGTGGTG
93121 GCATATGCCT ATAGTTCCAG CTACTATAGA GGCTGAGGTG GGAGGATTAC TTGAGCCCGG
93181 GAGGCAGAGG TTGCAGCAAG CCAAGATCGC GCGGCCACAG CCTGAGCGAC AGAATGAGT
93241 ATGCACCCAC GCCCTAAAAA AAAGCATGAC TCATTAAAAA AAAAAAATTT AGCCGGTCCG
93301 GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCGA GGCGGGCGGA TCACGAGGTC
93361 AGGAGATGGA GACCATCCTG CTTAACACGA TGAAACCCCG TCTCTACTAA AAATACAAAA
93421 TAATTAGCTG GGCGTGATGG TGGGCGCCTG TAGTCCAGC TACTCGGGAG GCTGAGGCAG
93481 GAGAATGGCG TGAACGCGGG AGGCGGAGCT TGCAGTGAGC CGAGATCGCG CCACGGCACT
93541 CCAGCCTGGG TGACAGAGCG AGACTCCGTC TCAAAAAAAA AAAAAAATAA AAAATTAAAA
93601 AAATATGAAG TTTTGAAGCA GAAATTATTT TGTCGTATGT TCTTTCATAA ATTTTTTGCC
93661 TGCTGCTCTT CTTCTTTTGT TACAGAACTC CAACACTTAC CCAAGGTAG CTGTTGGGTC
93721 AGGGTTTCTG TACTATAGTC CTTCTGTGG TGCCAGAAA TATGTTACAG GAAAGAGGTC
93781 CCCATCCAGA CCCCAAGAGA GGGTCTCTGG ATCCCGCGCA AGAAAGAGTT CAGGGTGAGT
93841 CCGCAGTGCA AAGTAAATGC AAGTTTACTA AGAAAGTAAA GTGGTGAAAC GACAACACT
93901 CCATAGACAG AGCAGGACAT TCCCGAAAGT AAGAGGAGGA AGGCATCCAC CCTAGGTACA

Figure 9 (Page 29 of 74)

SUBSTITUTE SHEET (RULE 26)

118/162

93961 ATACTTGAT ATATGGGGAG ATGTGCTCTG CTACAAGTTT GTGATAAAGG ATTAATTTTC
94021 TTAGTTACTA TATTTTGCAA GAATCAACAT TATTATCTTT AAACAAAATT AAGAATGCCT
94081 TTGTTCTCCA GATATAGGGA TATCTGGACA CTCCTAAGTC TGAGTCTGTT TAGTAAACAT
94141 TATTTATTTG TTCCCTTAAC CGTAAACATC TAGAAGCTAG GAATGACTGA CTTTCTGGGA
94201 ATGCAGCCCA GAAAGTCTCA GCCTCATTTT CCTAGCCCTC ACTCAAAATG GAGTTACTCT
94261 GGTTCAGTA ACTCTGACAC TTTTCTTCTC TTTTCTTCTT CTTTCTTCTT TCCTTTATTT
94321 TTTATTTTTT ATTTTGTAAA TAAGAAATCA AGAATACTTG ATGTTTCATC TAAACAATA
94381 CCCATAATTG ATAAGCCAAA ACAAACCTT AGGTCTTCTA ACTCAAAATG AGGATGTTTT
94441 GCTGTCTCTG CTGATACTCG GCTGATCGTT AATAGGTAAT TAACAAACAA GCCTTGCTAT
94501 GTCCCCCTCA GTTTATTACC ATTAGATCAT ATGCCTACTG TCAATCATAT TAATCCACAA
94561 CTATGCATTT CACAAAACCT GCCATAAAAA TTCACAGGTT TCCCGCTTCC CTCGAGTTTT
94621 CATTTCCGAA GGGTCCCATG TAATATAAAA CTTATATTAA ATACATTTGT ATGCTTTTCT
94681 CTTGCTAATC TTTTTTTTTG TTTTTTGAGA CTGAGCCTTG CTCTGTCACC CAGGCTGGAG
94741 TGCAATGGCG CGATCTCGGC TCACTGCAAC CTCGCTTCC CAGGTTCAAG CGATTCTACT
94801 GCCTCGCCCT CCCGAGTAGC TGGGACCACA GATACGTGCC ACCATGCCCC GCTAATTTTT
94861 GTATTTTTAG TAGAGACAGG GTTTCACCGT GTTGGCCAGG ATGTTCTCAA TCTCCTTACC
94921 TCGTGATCCG CCCGCTCGT CCTGCCAAG TGCTCGGATT ACAGACGTGA GCCACTGCAC
94981 CCGACCAATC TGTCTTTTTG TAGAGGGGCC TCAAGCATGA ACTTACTGAT GGGTGAGAAA
95041 AACAGAATT TCTTTTCCCC TACAATATAA ACATTAATTG TAATGTTATC ATTCAGGACA
95101 TTTTGGTGAC CAATCTTACA GAAATTTTAT CTTGTGCAAG TCTATGCAA CCAATATGTA
95161 AATCTTCTAT AAGTGAGATT GTATTTCACT TTTCTAGTAT CCTTTTAAAT TAATAAAGA
95221 GATTCTAATG ATTATTTTCA TTACTGCATT TCATTGTAGG GAAGTAGATA ATTGCCCTTT
95281 ATTCCTGAC CTTGCTTTT TAAAAATTA AACCATGTTA CCATGAAAAT GCTTTTCAGT
95341 ATTTCTCTAC ACACAAGATT GCTGTAAGGG CAAAAATAGA GATAGGAATC ATGCATCCAT
95401 TGATATACAT ATTTTGATT TTAATACATG TTACCAAGTT GCCTCCTGAA GGTCTGTTTA
95461 CACTCTCACC AACAGGGTGT TTTTCTCGA CTTCCACAAA TGCTCTTGAA CAGTGGGTGT
95521 GTTAGTCTGT TCAAATTGCC GACATGAACA ATTAAATCTC ATTGTTGTTT TTATTTTAA
95581 GACAATTATT GTTTGAGACT GCACATTTTG ATAATAACAT TTCTTCTATT ATGGTTTGAT
95641 TACTCATGAT TCTTGCCCAT TTTCTTTTGG GATGTTGCCT TATGTACATT ATTTTAAATA
95701 GATAGCTCCA TGTATTAAAA GATTATTAAG TTTGAGGGCT TATGATATGT CAGTTACATT
95761 TCTAAGATT TTTTTTTTTT TTTTTTGAGA CGGAGTTTCA CACTGTGTC CCAGGCTGGA
95821 GTGCAATGGT GCGATCTCGG CTCACCGCAA CCTCCGCTC CAGGGTTCAA GCAATCTCTC
95881 TGCCTCAGCC TCCCCAGTAA TTGGGACTAC TGGCAAGCGC CACCACGCTT GGCTAATTTT
95941 GTATTTTAT TAGAGATGAG GTTCTCCAT GTTGGTCAGA CTGGTCTCGA ACTGCCGACC
96001 TTGGCTTAAA AATCTACATT CTTTTTTTAA TTATAAAACT ACCACATCCC CCAAAAACAT
96061 TACTAAGGAA TTGAGGCTGC AGTTTAAAG GCTGATATT AGGATCTATC TCCGGAGAAG
96121 TGAGACCTGG TAATATAAGC ATTTTCAAAA TGAACCTTTG GGCCAGGTGA GGTGTGTCAT
96181 GCCTGTAATC CCAGCACTTT GGGAGACCTA GTCAGGCAGA TCACTTGAGC TCACAATTCTG
96241 AGACCAGCCT GAGCAACATG GCGAAATCCA GTCTCTACAA AAAATTAGCA GGGCGTGGTG
96301 GCATATGCCT ATAGTTCCAG CTACTATAGA GGCTGAGGTG GGAGGATTAC TTGAGCCCGG
96361 GAGGCAGAGG TTGCAGCAAG CCAAGATCGC GCCGCCACAG CCTGAGCGAC AGAATGAGAT
96421 ATGCACCCAC GCCCTAAAAA AAAGCATGAC TCATTAAAAA AAAAAAATT AGCCGGTCTGC
96481 GGTGGCTCAC GCCTGTAATC CCGACCTTT GGGAGGCCGA GCGGGGCGGA TCACGAGGTC
96541 AGGAGATGGA GACCATCCTG CTTAACACGA TGAAACCCCG TCTCTACTAA AAATACAAAA
96601 TAATTAGCTG GCGGTGATGG TGGGCGCCTG TAGTCCCAGC TACTCGGGAG GCTGAGGCAG
96661 GAGAATGGCG TGAACGCGGG AGGCGGAGCT TGCAGTGAGC CGAGATCGCG CCACGGCACT
96721 CCAGCCTGGG TGACAGAGCG AGACTCCGTC TCAAAAAAAA AAAAAAATA AAAATTAAAA
96781 AAATATGAAG TTTTGAAGCA GAAATTATTT TGTCGTATGT TCTTTCATAA ATTTTGTGCC
96841 TGCCTGCCTT CTTCTTTTGT TACAGAACTC CAACACTTAC CCAAGGTTAG CTGTTGGGTC
96901 AGGGTTTCTG TACTATAGTC CCTTCTGTGG TGGCCAGAAA TATGTTACAG GAAAGAGGTC
96961 CCCATCCAGA CCCCAGAGA GGGTTCTTGG ATCCCGCGCA AGAAAGAGTT CAGGGTGAGT
97021 CCGCAGTGCA AAGTAAATGC AAGTTTACTA AGAAAGTAAA GTGGTGAAC GACAACACT
97081 CCATAGACAG AGCAGGACAT TCCCGAAAGT AAGAGGAGGA AGGCATCCAC CTAGGTACA
97141 ATACTTGAT ATATGGGGAG ATGTGCTCTG CTACAAGTTT GTGATAAAGG ATTAATTTTC

Figure 9 (Page 30 of 74)

SUBSTITUTE SHEET (RULE 26)

119/162

97201 TTAGTTACTA TATTTTGCAA GAATCAACAT TATTATCTTT AAACAAAATT AAGAATGCCT
97261 TTGTTCTCCA GATATAGGGA TATCTGGACA CTCCTAAGTC TGAGTCTGTT TAGTAAACAT
97321 TATTTATTTG TTCCCTTAAC CGTAAACATC TAGAAGCTAG GAATGACTGA CTTTCTGGGA
97381 ATGCAGCCCA GAAAGTCTCA GCCTCATTTT CCTAGCCCTC ACTCAAAATG GAGTTACTCT
97441 GGTTC AAGTA ACTCTGACAC TTTTCTTCTC TTTTTTCTT CTTTTTCTT TCCTTTATTT
97501 TTTATTTTTT ATTTTGTAAA TAAGAAATCA AGAATACTTG ATGTTTCATC TAAAACAATA
97561 CCCATAATTG ATAAGCCAAA ACAAAAACCT AGGTCTTCTA ACTCAAAATC AGGATGTTTT
97621 GCTGTCTCTG CTGATACTCG GCTGATCGTT AATAGGTAAT TAACAAACAA GCCTTGCTAT
97681 GTCCCCCTCA GTTTATTACC ATTAGATCAT ATGCCTACTG TCAATCATAT TAATCCACAA
97741 CTATGCATTT CACAAAACCT GCCATAAAAA TTCACAGGTT TCCCGCTTCC CTCGAGTTTT
97801 CATTTCCGAA GGGTCCCATG TAATATAAAA CTTATATTAA ATACATTTGT ATGCTTTTCT
97861 CTTGCTAATC TTTTTTTTTG TTTTGTGAGA CTGAGCCTTG CTCTGTCACC CAGGCTGGAG
97921 TGCAATGGCG CGATCTCGGC TCACTGCAAC CTCCGCTTCC CAGGTTCAAG CGATTCTACT
97981 GCCTCGCCCT CCCGAGTAGC TGGGACCACA GATACGTGCC ACCATGCCCC GCTAATTTTT
98041 GTATTTTTAG TAGAGACAGG GTTTCACCGT GTTGGCCAGG ATGTTCTCAA TCTCCTTACC
98101 TCGTGATCCG CCCGCTCGT CCTGCCAAG TGCTCGGATT ACAGACGTGA GCCACTGCAC
98161 CCGACCAATC TGTCTTTTTG TAGAGGGGCC TCAAGCATGA ACTTACTGAT GGGTGAGAAA
98221 AACAGAATTT TCTTTTCCCC TACAATATAA ACATTAATTG TAATGTTATC ATTCAGGACA
98281 TTTTGGTGAC CAATCTTACA GAAATTTTAT CTTGTGCAAG TCTATGCAA CCAATATGTA
98341 AATCTTCTAT AAGTGAGATT GTATTCACCT TTTCTAGTAT CCTTTTAAAT TAATAAAAGA
98401 GATTCTAATG ATTATTTTCA TTAGTGCATT TCATTGTAGG GAAGTAGATA ATTGCCCTTT
98461 ATTCAGTGAC CTTGCTTTT TAAAAATTTA AACCATGTTA CCATGAAAAT GCTTTTCAGT
98521 ATTTCTCTAC ACACAAGATT GCTGTAAGGG CAAAAATAGA GATAGGAATC ATGCATCCAT
98581 TGATATACAT ATTTTGATTT TTAATACATG TTACCAAGTT GCCTCCTGAA GGTCTGTTTA
98641 CACTCTCACC AACAGGGTGT TTTTCTCTGA CTTCACAAA TGCTCTTGAA CAGTGGGTGT
98701 GTTAGTCTGT TCAAATTGCC GACATGAACA ATTAAATCTC ATTGTGTTT TTTATTTTAA
98761 GACAATTATT GTTTGAGACT GCACATTTTG ATAATAACAT TTCTTCTATT ATGGTTTGAT
98821 TACTCATGAT TCTTGCCCAT TTTCTTTTGG GATGTTGCCT TATGTACATT ATTTTAAATA
98881 GATAGCTCCA TGTATTAAAA GATTATTAAG TTTGAGGGCT TATGATATGT CAGTTACATT
98941 TCTAAGATTT TTTTTTTTTT TTTTGTGAGA CGGAGTTTCA CACTTGTTGC CCAGGCTGGA
99001 GTGCAATGGT GCGATCTCGG CTCACCGCAA CCTCCGCTC CAGGGTTCAA GCAATTCTCC
99061 TGCCTCAGCC TCCCCAGTAA TTGGGACTAC TGGCAAGCGC CACCACGCCT GGCTAATTTT
99121 GTATTTTTAT TAGAGATGAG GTTCTCTCAT GTTGGTCAGA CTGGTCTCGA ACTGCCGACC
99181 TCAGGTGATC CACCCGCTC GGCTCCTCAA AGTGCTGGGA TTACAGGTAT GAGCCACTGG
99241 GCCCGGCCAC ATTTCTAAAT TCTTTATAAG TATAAATTCA TTCAATCTTC ACCAAAACCTC
99301 AATGAAGTGT GAGTACTATT ATTATCATTG TTTTACAGAT CAAAACAAGT AATACAGTCA
99361 CTTACTGAGT TCTATACACC TGGTAATTTT TTTGTTTCGT TGTCTATCA ATTATTGGGG
99421 AAGGGGTGTT GAAATCTCTA CTTTAAATC ATGTATGTGT CTATTTCTCC TTTGCTTCT
99481 ATCAGGTTTT GCTACACATA TTTGTCAGTT CTGTTATTTG GTGCATATAC ATTTAGAATT
99541 GCTTGTTTTT CGTATTGGAT TGACCCTGTT ATCATTATGT AATATCCCTG TCTGTTCTTA
99601 GTAATTTTCT TTGCTCTGAA ATATACTTAT CTGATATATC ATCCAAAAGA CCACCAGGAT
99661 GGCTAAAGAG TAGAAAGGAG AGATTTACTG GCAATACTAA TTTGCAAGCC AGGAAGAGAT
99721 GGTCCCAGAA CCTGCCAAA TTAGTCTCTC TTTGGGGAGA AGGAGCAGGT TGGTTATTTT
99781 TATGCCTCAT AGGCTATATA TTACACAATA GAGTCATACA TATTTAGCAC GTTTGGGGGG
99841 ACAGCTATAT ATATTATGAG GGGTGCCAAG TGCATTCACA ATGGATAAAC ACGTGAATA
99901 TACCTCCCAT GTTCACTTCG AGGTTAAAT TTTGTTAAAA TGAGGTAGAA TTTAGGTCTT
99961 TACATCACAA GGTGAACTAT AGGAACAAAG TTTACGTGCT GCCTCTAGCA GCTGGCTGAA
100021 AATGGCTTAA GGTCTACAAT TACGTGTAAG AATAGAATGT GTGTCAAGGC GGTCTCTGT
100081 CCAATCAGAG TTGTAGTGGA CTGGACTGTA AATCAGAGTT AGGAGGGCTT CTGATAGCTC
100141 CTATAGTTAA GGAATTTAGC AAGTGTGAGT TTTTGGTAG TCTTTGGAAT TTAGGAATTT
100201 GCCATGCCAG CCAAGCCATG AATGCTCTAC CAGTAGGTAA CTTTGTGTTG TTAATCTTAG
100261 AGTCTGTCTT AGTTGGTATA GGGGCATCTA TTTTGGTCTT TCAGATCCCA GATATTATTA
100321 ATACAGATAC TCTTGTCAGT TTTGGCTGAT GTTTATATGG CTTATCTTTT TTGCAGCCTT
100381 TAATTTCAAC CTGCGTTATG TTTATATTTG AAGTGAGATT CTTGCAGACA GTGTACAGTT

Figure 9 (Page 31 of 74)

SUBSTITUTE SHEET (RULE 26)

100441	GTGTTTTTTT	TTTTTTTTGA	GATGGAATT	CACCTCTGTT	GTCCAGGCTG	GGGTGCAGTG
100501	GCACAGTCTC	AGCTCACTGC	AACCTCCGCC	TCTGGGTTTC	AAGGGATTCT	CCTGCCTCAG
100561	CCTCTTGAGC	AGCTGGGATT	GCAGCCATGC	GCCACCACAC	CCGGCTAATT	TTTGTATTTT
100621	TAGTAGAGAC	AGGATTCACC	ATGTTGCCCA	GGCTGGTCTC	GAACCTCTGA	CCTCAAGTGA
100681	TCCGCCAGCC	TCGGCCTACC	AAAGTGCTGG	GATTACAGGT	GTGAGACCTC	GCGCCCAGCC
100741	AAACTGTTTT	TTTATGGGTG	TATTTATACC	ACACACATTT	AATGCAATTA	TTGATATCTT
100801	AGGGCTTTAA	TTCATGAAGG	GTAGTGTGGG	AACCATAGTC	TCTTGGCCCA	CTAAATGTTT
100861	GCCAGAAATC	ACTGACAAGG	CAGATTGATT	AATAGGTGAA	AAGGCATTTT	ACCTATTGTT
100921	TAACGTGTCT	ATGTGGGAGC	ATTCAGAATT	AATTACCTAA	CTTCCCAATG	AGTTATAGAT
100981	GCTTATATAC	CATTTTTTAGA	TTCACGAAAG	AATTGGGGCT	TAGATTCTGG	TAAAACAGGT
101041	TATGGGAGGC	AAAAGAGGTT	TGGCTTGCAA	AGGTGGCCTT	GTTAGGTAGG	TGAAGCCTCC
101101	CTCAGAAAGA	ACAGATGGTA	AATGTTTCTT	TTATGATTTT	TAAAGTGCAG	ACTCTCAGTC
101161	TCTCCTGGAT	CTGGGAAAAG	GTATAGAAAAG	GTGAGGAGGC	ATGGCTGCAT	TAATGGAGAT
101221	TCTCTACAGA	TGTAAAATTT	TTCCCATTTA	AGGCAGCTTT	GCAAGCCCAT	TTCTGCCTGC
101281	TGGCCAAGCA	GCAGCCATTT	CAAAATATGT	CAAAGAAATA	TATTTTGGGG	TAAAATATTT
101341	TGATTTCTCT	TAGACTGGTG	GCCTTATAAG	AAAAGGAAGA	GACACCTGAG	CTGACACACA
101401	TACCTTGTCT	CTCTCAACAT	GTTATGATGC	AGTAAGAAGG	CCCTCACCAG	ATACTAATTC
101461	CATGCCCTTA	GCTTCCACAG	TTCTAGAACA	CATAGAAATA	AATTTCTTTT	CTTTAAAAGT
101521	TAGCCAGTCT	GTGGTATTCT	GTTATAGTAT	CACAAAATGG	ACTAAGTAAC	TATATTATGA
101581	TCATCTTACA	TGACTGATCC	CTCCTACATC	ATACACATAC	ACAGGCCACAT	TTTGGAACAT
101641	TGTTAGAGGT	TCCTCTACCC	AGTACAAATG	TACTACAAAT	TATATATGTA	TTTTTAAATT
101701	TTTGAGTATC	TTCAATAGTA	TATTTTCGTT	AACTTTGTGA	GTCAAATGT	CATTATAACA
101761	TGTATTCAAT	ATGCATAATT	ATTAGTCAGA	TGTTTACAT	TCTTCTTCA	TACTAAGTGA
101821	TATGGTTTGG	ATATTTGTCC	CCTCTAAATC	TCATGTTGAA	ATGTAATCTC	CAATGTTGGA
101881	AGTGAAGCCT	GGTGAAAGGT	TTTTGGATCG	TGAGGGTGAA	CCCCTCATGA	AGCGCACTCT
101941	TCAGGGTAAT	CAATGGGTTT	TCACTTTGAG	TTCACAAGAG	ATCTGGTTCT	TTAAAAGAGT
102001	GTGACACCTC	CCCCATCTCT	CTCGCTCAGC	TCTCACCATA	TGATATGCCT	ACTCCCTCTT
102061	CACCTTCCAC	CATGATTGGA	AGTTTCTCTG	GGACTTGCCA	GTAGCAGATG	CCTGCACCAC
102121	ACCTCCTGTA	CAGCCTGCAC	AACCGTGAGC	CAAAAAAAT	TACTTTTCTT	TATAAATTAG
102181	TCAGTTTCAG	GGATTCCCTT	ATAGTAATGC	AAGAACGAAC	TAACACACTA	AGTCTATTTT
102241	ATATTTACAG	AATAGCTCAA	TCTGAAGTAC	CCTTTTTTCAA	CTTCACAGTA	GCTACTTGTA
102301	GCTAGTGGGC	ACTGATTGCG	AGCGTGTTCA	AGGGTGAATT	GTATTATGCA	ATTAACAGAT
102361	TTTTTTTATT	GTTTTTCGCA	ACCACGAGGC	ATAGATTGTC	TTACTTCTCT	TGCTCCTGGT
102421	TTTGGAGTTG	TATTTGGGAA	ACAACCTTAT	TTCTCTTAT	ATTTATATGG	AATAAATAAC
102481	CCCCAATATT	TCCCTCCCCA	ATATCTGCCT	TTTGTATGTT	TTTTGAAGGC	AAGTGCCTAG
102541	AATTTACTGT	TTTTGAAGCA	CTTACTGAAA	GGATTGCCAT	CAAGTTGTTT	TGCTAATAGT
102601	ACATGCCAGG	CGCTTGTTGG	TTTGCTTAAT	TCAAGGTAA	TTGGATGAGA	AGAAGAGTTT
102661	TTCTCATCCA	TGGCTCAGTG	GAGTATAGAT	TACTGATATT	GTGACTGGAT	GTACTCCTGC
102721	TTTCTAGTCT	GAGTTTTTGA	AGCTACCCTT	AATCTTGGTT	TCAATTTTAT	CTAGCCCTGT
102781	ACATATCCAA	GGCTCTTTCC	AAAATGGTCT	ACGATTGTGT	TAGGAAGTTA	GAATAGCTGT
102841	ACTTTCTGAA	CCACGGTTCC	TGACATTTTC	TGGACTTCAA	ACACATCCAG	CATTTTATCG
102901	AAGTATTATT	CCTTCTTACT	TGGCTGGCTT	CTTCTTGCC	TTCAGGTCTG	AATTCAAATG
102961	ACATTCTCCT	GATGAAACTT	TCCACTCTTA	TTTCTATTCT	TTTTTCTTAT	CCCCTTTCTT
103021	TATTTTTTCT	CACAGCACTC	ATCATTATAT	TCTACTTTT	CATTATGTAT	TTACTTTATT
103081	GTGCACCTCC	CACTACAAGA	CAAGTAGCAC	CGTAAGGAAA	CAGGTGTGCT	GCTTTTTTAC
103141	TGCTATGCTC	CCTGCACCTA	GAACACTCTC	TGGCACTTAG	CAGGTTTTCA	GTAAATATAT
103201	GCTGAACATA	TAATGCTGGA	TATACATCTC	CCTCATGAAC	TCTCTAAATC	CTTCTAATTT
103261	ACATTGATCA	ATCTTCTTTT	CCATGTGCTT	TTGTATGATT	TATTGCTCAA	AATCTTTATT
103321	TTGTATGCAG	AACGTGCAC	GCTATTTAAT	CTTCAATGAT	GTAACTCTC	CCTTCTCTGA
103381	GTATAAATCT	TTCAGGGCAC	TATCTGAGAT	AACTTTTTAA	CATCTCCATC	ATGAATCTTG
103441	TACCTTTTCA	AAGAAAATGA	GCCAGTGATT	ACTGATGTTT	ACGGCTAATT	TTGAGGGTGA
103501	AGATCATTAT	AATTTTGAAA	AGGGAGATTG	AATATTGTGA	AGGGAAAGAT	AACATCAGAG
103561	TCAGAAGACT	TGGGAGAAGG	CAAAAAACAA	ACTAAAAATG	AGCACTTTTA	GTCTCTGAC
103621	AGTTTCTCTG	AATCAAATCC	ATAGTTCTGT	GACAGCGTTG	GCTTAGAAGC	AGATTTTTTT

SUBSTITUTE SHEET (RULE 26)

121/162

```

103681 TTTT TTTT TTTT TTGAAATGGA GTTTCGCTCT TGTCCAGGCT GGAGTGCAGT GGCACGATCT
103741 CGGCTCACTG CAACCTCTGT CTCCAGGGTT CAAGCGATTG TCCTGCTTCA GCCTATGGAG
103801 TAGCTGGGAT TACAGGCTCC CACAACCACG CCCAGCTAAT TTTTGTATT TTTAGTGAAG
103861 ACTGGGGTTT CACCATGTTG GCCAGGCTGG TTACGAACTC CTGTTCTCAA GTGATCTGCC
103921 CGCCTTGGCC TCCCAAAGTG TTGGGATTAC AGGCATCAGC CACCGTGCCC AGCCAGGAGC
103981 AGATTTTTTT ACATCATGT TTCTTTTCC TTCTGTCATC CTGTTTCACT ATAAGCAGAC
104041 CACAGATAGA AGTAGTAGAT ACCTCAGAAA TTCCTGGAAT AATTAAATCCA CGTTCATCTG
104101 TACTCCATCT GCTCCTATCT CATGGAATAT AAAAGGAAAA ACACCAAGAT TTCCCTAGGC
104161 AATCTGTCTT GATTTTAGGT TCCTCAACAG GAGAGCCAGA CAATGGCTGT AATAATATTG
104221 TCCCGGCCAA GGAAAAACTT CCCCTTTGCC CTCCCAAGGT TTATGGAAAA TTAGTGCCAA
104281 AACACAGATT AACTGGAGAA AAGGCATATA TATTATTTC ATCACAATTT TACAGGAGAT
104341 TTTAGAAATTA AGACTGAAAG ATACAGGGGA AATTGCCCAT TTTTATGCTT AGGTTCAACA
104401 AGATAAACAG CTGTATAGGG TACGATCTAA TGCTAACAGA CTGAGTGGGG AAGCCCCGCA
104461 AGGCTTGTCT GTCAAGATTG AATGGGGGGT CTTATGACCT ACAGGCAAAC AAGGTAGGTT
104521 AGGACAAGAC TCTCTTTTAG AATGGGGGGT CTTATGACCT ACAGGCAAAC AAGGTAGGTT
104581 AGAGTAATAT TTTTAGGTTT TATGGCTGGT TCTAGGGAAA AGGAGTTCTG GTTTGTATGG
104641 CCTACCTTGA GGAGGAATTC TGGTTTCTAT GGCTAGACTT TGGGGAGAAT GGGACTTACA
104701 GACAGGAAGG CAGAAGGTGG TCAGTGAAC ACTTTTATAA TCATAATCCC ATTTTGAGTA
104761 TTTCTGTGTT ATGGAATGTT TGTCTCTCA TTTCTGAAA GATTCCAGAG ACTCCTCATT
104821 CAGTGTGTG AAAAAGTTCA GGAATGCAA CTCAAAAATG TGCCACTTTG TTACGCTGAT
104881 TTCTTTGAAC TGAGGGCACC TAGGAAACAG TAAATTCAAG GAAGGGCTTT CGCTGAACTC
104941 TAATCAAAAA TTTGAAAATT AAAAAAATG TCAAAAAGGA ATTTAGTTGT TAAGATTCAC
105001 TTCCCTGGGG AATCTCATCA ACCAGAGAAG ATTAAGTGA TCACAGGAGA GGAGACTGGT
105061 GGTTAACACC ATCTAAACAG ACTTGTGAC AGCTGTGACC TATCTTTGA AACACCCATT
105121 TATTTTTTCT CAAAATCATA TACTCTCCCC TAAGTTGCCT ACATCCCCCT TCCTTCTCCC
105181 TTATGAATCA AGAGAGCTTA TAAGCTTCTA CAGTTCAGTG GGATTTGGGG TATTCGCTTT
105241 TCTTCCCTCC CACTCCCCCT CCCCTTTTTT TGTCTTTGAG ACACAGTCTT CTGGCTCTGT
105301 CGCCCACGCT GGAGTGTGGT GGCTCTATGT GAACTCACTG CAACCTCCTC CTCTCGGGTT
105361 CAAGCGATCC TCCCACCTCA GCTTCTCGAG TAACTGGAAC TACAGGCGTG CACTACCAAG
105421 CCCGGCTTTT TTTTTTCTT TTTCTCCCC GTTCTTTTT TGGTTATTTT ACTGGAGACA
105481 GGGTTTCTCC ATGTTGTCCA CGCTGGTCTC GAACGCCTGA CCCGCCGTCC TCGGCCTCCC
105541 AAAGTGTGG TATTACGGGC ATGAGCCACT GCGCCGATT TGAAGGACCT CTTAAATATC
105601 TATTTAGAAA TTGGTCGGAG TCCACTCCTT TCCAAAAACA TGAGTCACAA TCCGGGAAAA
105661 GCACGAGCGG CTGAAAGTCA AAATAACCAG AACAAAACCT CCACTCATGC TTAAAAAAGG
105721 TATTTTGACA AAATCCTAAT TCGGCCAATT ATTATTAGTA TTCAAGTCGA AGGCTCGTCA
105781 AGCCAGACTG GGGATTGGGT CAAACATAAA CCTTACACCA GACGGAAGGA TTACATGCAA
105841 ATGAAGGATG CAGATTCTGA TTTCCCATG GGTATTTGAC ATTAGCCAAT GGGAGAATTC
105901 CTCACAGCCT ACCTCCAGTC AGTATAAATA CTTCTCTGCC TTGCGTTCTA ATGTAGTTTC
105961 ATTACATTTT CTTGTGGCGA TTTTCCCTTC TTATCAGAAG TAGTTATGTC TGGTCGCGGC
106021 AAACAAGGCG GTAAAGCTCG CGCCAAGGCT AAGACTCGGT CTTCTCGTGC AGGTTTGCAG
106081 TTTCTGTGG GCCGAGTGCA CCGCTGCTC CGCAAAGGCA ACTACTCCGA GCGCGTGGG
106141 GCTGGCGCGC CGGTGTATCT CGCGGCGGTG CTTGAGTACC TGACCGCCGA GATCCTGGAG
106201 CTGGCGGGCA ATGCGGCCCG CGACAACAAG AAGACCCGCA TCATCCCGCG CCACCTGCAA
106261 TTGGCCATCC GCAATGACGA GGAGCTTAAT AAACCTTTTG GCGGTGTGAC CATCGCGCAG
106321 GGTGGCGTTT TGCCTAATAT TCAGGCGGTG CTGCTGCCTA AGAAACTGA GAGCCATCAT
106381 AAGGCCAAGG GAAAGTGAAG AGTTAACGCT TCATGCACTG CTGTTTTTCT GTCAGCAGAC
106441 AAAATCAGCC TAACAGCAAA GGCTCTTTTC AGAGCCACCT ACGACTTCCA TTAAATGAGC
106501 TGTTGTGCTT TGGATTATGC CGCCCATAAA GATGTTTTTG AGGTGTTTTT AATGGCTTTG
106561 AGTGTGGCAC TTTTAGTAAT TTGTCTGCA GAAATTAGAT CCATAGAAAC CTCAGGAATT
106621 CTAGGTATGT GGGAGAAAGT CCAATGCAGCA CAAAACATGT TTACAGGGGT GATTGCGGTT
106681 AAGTTTCACA CACAGCAGTT ACTACATTTT AGAGGAAGGA AATTATACCC ATGAGTGCAT
106741 TCCTAACTAT CTTGAATGGA AGTGTTAAAA CCCGCATGCC CCACACAAGT TTGAATATGT
106801 CATACATTT GCTGTAGCAA TTAATGGCAT ACACAATTGA GAGCACACAC ATTACCACTG
106861 AACATTTGAG TATGTATTTC CCAAATGAG CTTTTTTCCA GTTTGGGGAT GTTTTGCTTT

```

Figure 9 (Page 33 of 74)

SUBSTITUTE SHEET (RULE 26)

122/162

106921	GTTTTGGGGT	GGAGTCTCCC	TCTCGCCCAA	GCTGGAGTGC	AGCGGCGTGA	TAACAGCTCA
106981	CTGTAACCTC	GAACTCGGGC	TCAAGCGATC	CTCTTGACAG	CCTTCTGAGT	AGCTGGGATT
107041	ACAGGCGAGA	GCCGCCACGC	CCGGCTAAGA	GCATTTTCT	AATTGCCAC	ACTTCTTATG
107101	CGACACCCAG	AAAAATACAA	TTTTAAATAA	AGCGCATATG	CAAATTTCCC	TAATCGTCTC
107161	CAATATTCTC	TGATTTCTTT	TTTATATTTT	AACTAGAAAC	AATTGGAGGT	TTCCGCGTTG
107221	CTTTGTGTGG	TTGTAAATTT	TAAGACTTCA	GGAAACTTTT	CCAGTACAAG	ACTTGTCCAC
107281	AGTGGATATA	GCAGCTAAGG	GGTTAACAAA	ATGACGTCAG	AGTAGCTACG	GTAATGGGCA
107341	GGAGCCTCTC	TTAATCTGCA	ACCAGGCACA	GAGATGGACC	AATCCAAGAA	GGGCGCGGGG
107401	ATTTTTGAAT	TTTCTTGGGT	CCAATAGTTG	GTGGTCTGAC	TCTATAAAAG	AAGAGTAGCT
107461	CTTTCCTTTC	CTCCACAGAC	GTCTCTGCAG	GCAAGCTTTT	CTGTGGTTTT	GCCATGGCTC
107521	GTAATAAACA	GACAGCTCGG	AAATCCACCG	GCGGTAAAGC	GCCACGCAAG	CAGCTGGCTA
107581	CCAAGGCTGC	TCGCAAGAGC	GCGCCGGCTA	CCGGCGGCGT	GAAAAAGCCT	CACCGTTACC
107641	GCCCGGGCAC	TGTGGCTCTG	CGCGAGATCC	GCCGCTACCA	AAAGTCGACC	GAGTTGCTGA
107701	TTCGGAAGCT	GCCGTTCCAG	CGCCTGGTGC	GAGAAATCGC	CCAAGACTTC	AAGACCAGTC
107761	TTGCTTCCCA	GAGCTCTGCG	GTGATGGCGC	TGCAGGAGGC	TTGTGAGGCC	TACTTGGTAG
107821	GGCTCTTTGA	GGACACAAAC	CTTTGCGCCA	TCCATGCTAA	GCGAGTGACT	ATTATGCCCA
107881	AAGACATCCA	GCTCGCTCGC	CGCATTCGCG	GAGAAAGAGC	GTAATGTAA	AGTTACTTTT
107941	TCATCAGTCT	TAAAACCCAA	AGGCTCTTTT	CAGAGCCACC	CACCTATTCC	AACGAAAGTA
108001	GCTGTGATAA	TTTTTTGTTG	TCTTAACAGA	ACAAATTTCT	AAGGACCCCC	CCGGAAGACA
108061	TTAGACTATG	GTCTTAAAGT	TGATTAACAG	AAATAACGGT	TTGGTCAGTC	TTGCAGTGTA
108121	GGTTATTTCT	GACCTTATTA	AGGTGCTATT	TGGAGAGAAG	CTGTGTAAGT	CCACTATCAT
108181	TCAGGCCTCT	AGCTTGCTAT	GATTAGCATT	TGTTTAAACA	ACTTTGTAAG	AGTAAGGGAA
108241	AAATCTGGTA	AGTAGTTAAC	TGGCGCTTAC	TAGGCATTTT	TGCAAAGCTT	TGAAAAGATT
108301	AGAAAATTGT	GTCTTGCGAG	TTCCAGTGTC	TTCTCAAAA	TGCTTAGGAA	GATTTTCTCA
108361	GCTCAATACA	TAGTCCCCTA	GGTTTTCTCA	TATATTATAT	ATATATATAT	ATATATATAT
108421	ATATATATAT	ATATACTGTT	AAATTCATTT	GGCTGTAAAC	ATTAACCTGA	AATTTATTCT
108481	GGTGCAAAAT	GTGAGGCAGG	GATCTCATTT	GCTCTCATTT	TATCCATAGC	TAGCTACCCA
108541	CTTTAAATCT	GTCAGTCTGT	CGACCAAGCA	TAATTTAATC	CCTTATATAT	GAATTTTTAT
108601	ATGTGTGGCT	TTGCTTGTA	ATAGTCTATC	TGGTTGCATT	GCTTTGTCTC	CTCTAGGACT
108661	ATGCACCATG	ACATGCCACA	TTCTTTTTTT	CAGTACTTCT	TGCCTGTAGT	TATTAATAATC
108721	TAGAATTTAC	AAGTTTTTAAC	CATTTTCTTT	CTGTTGATCT	TGCTTTTCGG	TTTTGGAGGT
108781	TGGGGATTGA	GTAAGGGAAG	AAAATTTAGA	GGGATGGGAA	TACTGTACGC	AAACAAAAGT
108841	AATATTTACT	TTAAAATTTT	TATATTTTGT	ATTTTTTTAT	CATATAGCTT	TTACATCACA
108901	TTTTACAGAC	TAACCTTAGA	ACAACCACAG	AATGTCCAAC	ATTAATACTA	CTAATTTCAA
108961	AGACCTTGCC	TCACATTCTT	TTTACAATA	AATATTTTTT	ACACCTAACA	TTCTTTCTTG
109021	GCCTACATCT	AGAATGTAA	CTGATGTACC	ATACTAAAT	CGCCTGACCA	ACTGTCAACA
109081	ACAACAAATC	ACACACACAA	AAGATCAAAT	TTGAATTGCA	TCGTTTACTT	AAATTCATTT
109141	GTGTTCCAGC	TTTTAATAAG	GCAGTTTTTG	GTTTATAAAG	TAATATTTGC	ATTTTAAAAA
109201	TTATGAAAAT	GAATATGTCA	GTTTGTTTTA	TGATTGCTTT	TTCTTGACTC	TTATACAAGC
109261	GACTCTAACT	GGCATAGACA	TTTGTTATCC	ACAGACAGTA	TAGATATGTT	AGAGATGCCA
109321	ATGGACTTGG	TCTATGCCAA	GGTGACTACT	CACAAGCTCT	GGGCCCAGCT	GAAGGTCAAG
109381	TATTTTTTTT	CCAGTTATAG	ATGTGCTGGA	TCTGATGTAT	AGCGCTTGAC	TTTTTATATT
109441	TTCTTTATCT	GTAGGAAACA	AATGTGTTGG	AGGTACTGGG	TCTGACGAAT	AGCATAAAAG
109501	AATAAAGTTA	CATTACTGTC	TGAGGATCAG	ATGGACAGGG	GGTGGTAGCT	CAGTCCAGCT
109561	ATTTTCCACT	CCCTCACTTA	CATTCTTTGC	CCCCTCCTCA	ACAGAACAAG	GATTCTGCTG
109621	TAACCTCTCA	TTGACAGTTG	ATATTTAAAA	ATTAACGAAT	GGATGAAATT	CTCATTTGTG
109681	AAAGAAAATT	TATTGAGCAT	TTTGATTTTG	TGAGTAGTGC	AAACATTTTA	ATATTATATT
109741	AAGAATCTAT	TGTTTTGTAT	TAGAGGAGTA	ATTAAGGAGA	GATTGGAGAC	AAAAAGGGGG
109801	TGTTGTTTGC	AGAATATACC	ATCCAAAAAT	AGACCACTGT	GGGATCAGGA	TTCTTTTGAG
109861	CTAAAGGCAC	TTCAAAAACA	GCATTCAAGA	AGGGAATTCT	TCTAACTTTT	TCTTTCTGAA
109921	AACAGGAGAT	AAAAGTTCCA	ATGTGAAAAA	TGCTCTGCTT	GTACCAGGTG	AAAAGACATA
109981	TTCTTCAGCC	CAGAGGCATA	TTCGTGACAA	ACACAGCAGG	GAGTCATAGC	
110041	CGAGAGACTT	CTATACACAA	ACAAACCTTG	TTAAATAAT	CATATATTCC	TTTAATCTCC
110101	TCATATGGTT	TACTTTCCCA	CAATTGCCTC	TCTTTAACTT	AATGTGAAAG	CATTTAGCTT

Figure 9 (Page 34 of 74)

123/162

110161 TTGCCATTTC TTTGGGGCTT CACTTTTTTA TGAGGGTTCT CCTGTCCCAT AAAATTTACA
110221 TTAAATACAT TTGTATGCTT TCATTCTGCT AATCTGTTTT ATGGCAAATG AATTATCAGG
110281 TCCAGCTGGA GACCCTAACA GAGTAGAGGT AAAATTTTGC CTCCTTACAA GATAGAGATT
110341 GTGTGCATTA AATGTTGTTT GTTCCCAGTT GTTCAGTTTG TCAGGCCTCT GAGCCGAAGC
110401 TAAGCCATCA TATCCCTGTG GAACTGCACG TATGCCTCTA GATGGCCTGA AGTAACTGAA
110461 GAAACACAAA AGAAGTGAAA ATGCCCTGTT CCTGCCTTAA CTGATGACAT TACCTTGTGA
110521 AATTCCTTCT CCTGGCTCAT CCTGACTCAA AAGCTCCCCC ACTGAGCACC TTGTGACCCC
110581 CACCCCTGCC AGCCAGAGAA CAACCCCTT TGAAGTGAAT TTTCCACTAT CTACCCAAAT
110641 CTTATAAAAC GGACCCACCC CATCTCCCTT CGCTGACTCT TTTCCGACTC AGCCCGCCTG
110701 CACCCAGGTA GAATAAACAG CCTTGTGTGCT CACACAAACC CTGTTTGATG GTCTCTTCAC
110761 ACGGACGCGC CTGAAACAGT TTAACAGGGT TTTTCCTGCC CAGTCACAAC AAAGTGATGT
110821 TATGCTGCAG GCTGAAGTTT ACAGCTAATG CTGTTGAAGT CTAAAATCAG TTTTGGTTTG
110881 TTAGATTTGG GTGAGATGGC TAAGATTCTC AGAGAAAGAA GTCAAGTTTG GGGTGCATTT
110941 TTCAGACTTA AAAATTTAGC AGTAGCCCTT GCAGTTTTTC CAATAGAAGT GATTTACGAA
111001 TGTTTTTCAGG AAATTTAAAA CAACAGTGAG AAGCGTGAT GGAGAGTTGA ACTACACTCC
111061 AGACTTGGCT ATAGGAAAGC ACGAATGCTG CTATTGTATT GCACCTTGGA AAAGAGAACA
111121 AAGGAATATT TTCGGACAAT TTTAACATGT CACATATGAA AAGCTAAACG GAATCTGTCA
111181 ACACCTTGTA CGTTATTACA GGCTGTGATT TTAACAAAAC AATCCTTACT AATACATACA
111241 TAGTTGCTGC TAGCAATATA GTGTTGGGAG TAAAAACACG AAAATGAGAG TTCAGGACAA
111301 TATCCCAACT CTGAGCAGAT TTTTAAAGT AGTAACATCT AAAATTAAC CATATTATGT
111361 AATATTTATT TCTTTCCAC AGTCTCTTCT CATGCCTCGT TCACATTAGC TAATTAAGG
111421 TCCCTGAGT ATCATCATAA CCCGATTTAC AGATGAAGGC ACGGTTGCAA TGAGCTATCA
111481 CCCTCTTCTG AATGAGACAG TACAGTGTA AGGATAGCAA AACTCCACTC CCATCCTCTT
111541 AGGGCTCTGG CTGGACCAGC AAATTAATTT AATGTAAAT GGATTAACAG GAGAAAGGTA
111601 TATGCATTTA TTTAACACAG GTTTACGTG ACACAGGTGC TCTCATAAGG TAATGAAAGC
111661 CCAAAAAAG CAGTTAGCTA CTTATATAAT GAATTGGACA ATTAGTAAAA TGTAAAAATG
111721 CGCTAAAGCA AAGGGATTTA GGCTAGAATA TATAACTGTG TAGAGAAGCG CCCAGCAAGG
111781 GCTAGTGCAA GGTGTGTACA GAATTCTCTT GGCCTCAGCC TCCTATCCTT GAGAAGAATG
111841 TTGCTTTTTT TAAACTACAG TGAGAACATC TTTCATATGA GAATTTACCC TACTGCTTCT
111901 AAGAAACAGG TCAGCTTTCA AGAAAACATA AGGCCAGAGT GATCTTTTCA CGCCTGCTCT
111961 TTTAAGTACC TTTGAATAGT CAATATGTCT TCAAGCACTT GAAAGACTTA AAAAGTTTAC
112021 CACTCCGGCA TATTAGTGAA AGCCCTTAAT ATAAGCCCTT ATTAAATTC TCAGTCGAGG
112081 GTATAAATC AGATTCAAAT AGTAGTGCG TAAACGGGAG GGAAAACTA AAGGGATTAA
112141 AAAGTGAAAC TATTGTGTTT TCCTCGCAG TCCTTAGGTC ACTGCCCTC GAGGGGCGGA
112201 GCAAAAAGTG AGGCAGCAAC GCCTCCTTAT CCTCGCTCCC GCTTTCAGTT CTCAATAAGG
112261 TCCGATGTTT GTGTATAAAT GCTCGTGCTG TGCTTTCTTT TCGCGTACCT GGTTTTGTGTT
112321 GTCAGCTGGT TAGACATGTC TGGTCGCGGC AAAGGCGGTA AAGGTTTGGG TAAGGGAGGT
112381 GCTAAGCGTC ACCGAAAAGT GCTGCGGGAT AACATCCAAG GCATCACCAG ACCGGCCATT
112441 CGGCGCCTTG CTAGGCGTGG TGGGGTTAAG CGAATTTCCG GTTTGATTGA TGAGGAGACT
112501 CGTGGCGTTC TCAAGGTGTT TCTGGAGAAC GTGATCCGGG ACGCCGTGAC CTACACGGAG
112561 CACGCCAAGC GCAAGACTGT CACTGCCATG GATGTGGTTT ACGCGCTCAA GCGTCAAGGA
112621 CGCACTCTGT ACGGCTTCGG CGGTTAATCT TTTCTGTCAGT TTTCTTCCAA TGGCCCTTTT
112681 TAGGGCCGCC CACTCCCTCT CAGAAAAGAG TGTGATTGTA TTCTTTCCGA TGGTAACATC
112741 TCAATGGCTT TACTCGGCTA TTCTGCCTAG TATGTAGAAC TATTATAAAC CAGTTGGGAG
112801 AGACCAGGTT GTTTGGTCTG AGTGGCTGCT AAAGCAGAAA TCAGCTAAGT AAACGAGGTC
112861 TCCGAGATAA GTGAGCTATA AACTTCAATG CTATAGTTTT GACATGTCAA GCAACTTAAC
112921 GTGCAGCGCG AGTCCGATAA ATGAGTAGCT CAGCTTTTTA GTTTTAAAAA CGAGTTGTGC
112981 GTTATTTGTA CGAGAGCCTA AGATGCTAGC TGCCTGGAAC TGAGTAGGTG GATTAAAAATG
113041 GGTGTCAGGT CTGTTTTCCC AGGCGTATCT GACTTAACGT CAGCAAAGC TGTACTTTTA
113101 GCTTCCCTGG TAACACCTGC CGTCCTTAAC CGCCCTTGC CGGTAGCGCC AGAAGCCTTT
113161 ACTTCCATTT CTAGTTGAGC TTGGCGTCTT GCTGAGTGAC GTCACCTCCC CTTCTGTGG
113221 AGTAGGACTG GCGGTTAAAG CTGCTTGTCT ATTTTCAGTC CTCAGGCTGG AGGCTCCCTT
113281 AAGCAGGCTG CCTACGCGAG TCGTAAATTC CCACTTAGTA GACTAAGGGA GTCTGTTTTA
113341 TAAATAAGGA CTCAAATTTT TTCTGACTCC GAGGTCCGTG GCAGCAGCTA TAAGATGGAA

Figure 9 (Page 35 of 74)

SUBSTITUTE SHEET (RULE 26)

124/162

113401 GCCCCCTCTG ATGTAAGATT CTCAGATGAC TTGCATCTTC ACTGTACCTG TCAACCCAAT
113461 AGTCTTCTAT TCCTGCCTTA AATTGTAAAT TCCAAACTG ATTTAATTGT GAAAGTTTCA
113521 AACTGTACGA CCTAGGAAGT GTCAAAGTTA GGTGACCAGA TTTTGTAGAAG TCAGCCAAAT
113581 ATTGAGCATC TTTGATTAG TAACAAATAT ATTGATGGCT ACTTCAGCAA AAAAAATCAA
113641 CTTTGTTTTC TGGTTACTTT GCTAACAAAGC TTCTCCTGAC AGGAGGATAT AGTGAATAGG
113701 CAGTTGAATA AGTGAGTTCG GGTGAGAGGT CTGAGCTGGA GATAAAAATG TGTGAGTCAT
113761 CAGCAGATAA ATAAATGCTG AGACCAGATG AGATGGCTAA AAACGTAAAC ATAATGTAGT
113821 GCAGCATTGT TTGTAATAGT AAATGAGTGG CCACTGTAAA GTTTTTCATCA GAAAGGACTA
113881 GAGTGATCTA TACATCCATA AAATAGAGTA TTTCTCTACA CAGCCCTACT AAAGAATGAG
113941 AAAGCTGTAC TCCACTACAT ACTCTGGTGT ACTCTGGCTC AGTTCCTGGA CTCCTCTTTT
114001 CTTGGCTAAC TCAACTGGCC TCACCACTTA CATGCTCTGT GCTCTGTCAA ATAGTTTGT
114061 CAACAGAACA CCACGGCCTA GCTGTAAGTG CCACGTAAAC TTCTAGCAAT GCCAAAGCCT
114121 GTGATAGTGG CAGCTTCGGG CTGTTTCTCA TTCCCGGGAT GCCTAACCCAC CTCTCCAAAT
114181 TCTATCAGTT TGCTTCACC CACTTCAAGC TTCAGAACGA AACATAGAGC TTAAGAAATA
114241 TAGGCCCGGC AAGGTGGCTC ACGCCTGTAA TCCCGGCACT TTGGAAAGCT GAGCCTGGTG
114301 GATCACCTGG GGTGAGGGGT TCGAGACCAG CCTGGCCAAT ATTGTGAAAC CCCGCTCTTA
114361 CTAAAAAATA AAAAAAATTA GCTGGGCATG GTTGCGGGCG ACTGTAATCC AAGCTCTCG
114421 GGAGGGTGAG ACAGGAGAAT AGCTTGAAC CGGGAGGCAG AAGTTGCAGT GAGTTGAGAT
114481 CGCGCTATTA CACTTAGGCC TGGGAGACAA GAGTGAACT GTGTCTCTAA ATAAGTGT
114541 GCAATTATAA ACCATCTCCC TGACCTTAAA TCTCTAGACT CATATACAAC TGCATATTTG
114601 ATGTATCTAA TTGAATAATG GGCATCTCGA ACTGTCCAA AATATGTTTA TACGTAAACA
114661 CCAAGTCTGT TCTTCCTCTG ATATTGTCA TGTCATCAA TAGAACTCCA TTCTTCAAGC
114721 AGCTTGGGCC AGGAATTGTG CAATATTGTT TGTCCTGAGC TTCTTACAAC TTTCACCCAA
114781 TGCACTCAGC TCTGTTGAAA ATCAATCAGA ATACCTTTCA TTGTTTCTT TGCTGCTTCT
114841 CTAGGAGCAA GCTGCCATGG CGTTTGTCT GAATGACCAC AGTGACCCCA AACTGGTCTT
114901 TGTTTTCACT TTTAATCCCC CTGTCATACA GTTTTCTCT ATCCAGCATC AACAGTGATC
114961 CTTTTTGAAG GTATTATGTC CACTGTCTGC TGAAAAGATT CCACTGGCTT TCCATCACCT
115021 TCATAATAAA AACCAGCATC CTTATCATAG CCTACAAGTA AGATGACCAA CCATTACAGT
115081 TTGCCTGACT CTCAGGGGTT TCTCAGGGTG TAAGACTTAC AGTGCTGAAA CTTAGAAAGT
115141 TCCAAGCAAA CTAGGATGAG CTGCTCAACC TACTAGATCT GTACTCTGGC TACCCTCTGA
115201 CCTCATTCTC TTCGCAGTTC TTTCTCTTCA CTGACCTTGC TGTTTCTGGA ATGGACCAAG
115261 CATTTCCAGC ATCAGCACCT TTATATCTAT TCTTTCTCCC TAGAAGGGTC TTGTCTGGA
115321 TATCTGAATG GCTCTAGATC TCATTTCATT CAAGCCTCTC CTCAAATACC AACCTTAAGA
115381 AAGAGACCTC CCATAATCAT CCCTGTAAA ATAAGCTTTT CTGCTCATTT AGCATATATA
115441 TATATAGTTG ACTATCCTCA ATAGCATATA TATATAACAT TCCCCACCT AGAATTATAT
115501 ATGTAATAAT ATATTTAACA AAAAAATCAT ATAAGTAGAT ATATTTTATT TTGTGTTTGT
115561 TCTCTCTCCC CCAACTGGAA TATATTTTTT GAAGGTAGGG ACTTTGTTTT GTCCCAGAAG
115621 TATCCCTAGC ACCTTGAACA GGGCTGACGT TTAACAGGTA GTTTATGGAG GTTTGTGAA
115681 TGAAAGGATG TGTGAATTTT CTATGTAAGT CTCCAGGCTC TCCACTAAGC CCACCAGAAT
115741 GCTAACACAA TCAATTCCCC ATCTCATTCC TTGACCTGCC ACTGCCTGAA GCAATCAGCG
115801 TGCAGTTTCT CTTTAGAAAA TCTGGGGGAT AGTCTAGGGG TTGCAAATTA AGCAACATTA
115861 TCTTTGTTCT GAACAAGGAC TGCAATAGTG TTAGGACTGA AGAAGGCCCA AGGTGGTGGT
115921 GGGTATGCCT AAGATGAGTA TGACATATCA GCAATGCTAT GAACATAGCA ATGCTATGAA
115981 AGGCCAGGCA AAACGTAACA GGAGCTAGTC GTGGCTTATT GTTACAACGA CTATACCTCC
116041 CATATGGGTA ATCGATATCC ACACACCCCT CTACATTGAC TCTGGAATTC AGGAAAGGGA
116101 ATTAATAATT TCTAATTAT GTACCCCAAT GATTTCAACA ATATCTGGCA TATGAGATCA
116161 ATAAATATCT TTAATAATACC AACTAAGAAA GACATAAAAT GACCCACCTT CCATACCAGG
116221 CTCATTTTTG CTCCTCTGAT TCCTGAAACT ATCCAGAATG CAGCTATGAA TTCTCTCCAT
116281 TGTCAAGTTT AAATTAAGCC AAGCTGGGTA CTTGTGTAAT TCCTCAAGAA ATCCTGGATG
116341 AAAAGTGTCA GGTGAAAAAC AGGACCTCAA AATAAGAGA CATCCATCAC TGAAGCTAAC
116401 ATCGTGAGGC TGAATCAGT CCTATAACAA TGGTACCAA AAGAGCACAA TGAGAGGCAT
116461 TTGTGAATAT TTAATCAGAT GAGATAAGA TATTTCCCTA TCAGCTAAC TGAAGTTTCA
116521 ATCCCTTTTC CAGCTGAGTT CTGAAGCTAG ATGTACTTAA CTGGAACACA TAACTGCATC
116581 AGGAACATCC TTTAAACTA TGGCTACAAT GGCTTGACTG GACAAACCCC AGGCTTCCAG

Figure 9 (Page 36 of 74)

SUBSTITUTE SHEET (RULE 26)

125/162

```

116641 GTTTAGCACA GGTGGCCCTT CACAGACCAA CATTGCCTAT GCTACCAACC TCATGTCCTA
116701 CCACCCTGCT TGCATCATTT CTCTCTCTGC ATATATAAAA ATATATGTGT ATGTATATAA
116761 TCAGCTTTAT TGATATTTAA TATACCACAA AATTGCCCCA CTTTAGGTAC AGTTCAATGA
116821 ATTTTACCGT GTTTTCTTAG TTGTACAACC ATCATCACAA TTTAATTTCTG GAATATTTCT
116881 ATCACCCAAA TTTCCATTTT TCGGTAAAGG GGGAAAAAAA AAGGTTAACT GCTGAAGGCC
116941 GCGGTAACAC TGAAAAAGGT GCCTTTTCTC TCTAAAACAG ATTTTAATCT CCCCTGAATT
117001 TAGTGTCTTG GGTATTCCAG GAGTCTGAAT AGGGTTTCAA TTTTCAGGGT CTTTTTAATA
117061 GAGTAAACT GTATTGGTGG CGATAAATTT AGTATTGCTC TCAGTACATG ATTGAGGGAT
117121 ACTTAAATGT CTCTGTGATT TTATTTTATA ATCGCTAAAA GATGGTTTTT TTTTTCCTA
117181 AAACAGGGTT TTTGTTTTTT CTCAATAAGC TTCTTAGCTT CCCCTCCGGC TCCCTGGCTT
117241 GCCTCAGGAA ATATTAGCTC ATCAGTTCTG ATTGGTTGAC AGCTACGAAT GGCCCTCATT
117301 GATTGGGCAG CGCTTCTTTG TCCCTTGGAA ACTAATACAA ATTTTAAACA CTACTTTTTT
117361 TCCACTCTTT CTTCAGAGTT GGAATATCGT TGCTCCCCTA CCCATATGTA GTGAGTGGAG
117421 GGCAAACCTG GAGTTCCCTT AATCTTCTCT TTTTAGGATG TCAGCTCAGT ATCATTCATC
117481 TTAATTACAC ATTGAGCTTC TTGACTTAAT GGATACAGCT CTTCTTTTGT TTAGTTGGGC
117541 GGCCCTGAAA AGGGCCTTTG GTTCAGAAAT GCAAGCTGTG GAGAAATCAG CAACCTTAAC
117601 CGCCAAAGCC ATAAAGGGTG CGTCCCTGGC GCTTAAGCGC GTAGACCACG TCCATGGCAG
117661 TGACTGTCTT GCGCTTGGCG TGCTCCGTAT AGGTGACAGC GTCACGGATC ACGTTCTCCA
117721 AAAACACCTT GAGCACCCCG CGAGTCTCCT CGTAGATCAG ACCAGAGATC CGCTTCACAC
117781 CGCCACGCCG GGCCAGACGC CGGATGGCCG GCTTGGTGAT GCCCTGGATG TTGTCACGCA
117841 ACACCTTGCG GTGGCGCTTG GCACCCCTCT TACCCAAACC CTTCCCGCCC TTACCACGTC
117901 CAGACATGAC TTCCCAAGAA GTGAACCAAG AGCAAGTGAG AGAATAGGAA ACCGATCTTT
117961 ATATATCTAC GTTACCCCTG CCCCCACCTC CAGCGGACAC AGAGACTGAA AAGCGCGCAG
118021 GCGGAAATG TGACGCCTAC AGTCCGCTCC TTAAACCCCT CCTCCAAGCC CCAGGAAATG
118081 GCGGGAGCAG CGATTGGGGG AGGGTGGGA GATGAGGGTG GGACCAAGCA GGCTTGACCA
118141 ATGGCCTTTA TTTTCTTAAC AGAGCTACAG GCTTTGAGGA ACTGGGTTAA GAATTAAATG
118201 TAAACCCATT CTGACTCCAG AATTATTTTA AGTCGAACTT TTTTTTTAAC CGAATCTCTC
118261 TGTCGCCCAG ACTGGAGTAC ATTAGAGCCA TCTCGATTCA CTGAAACCTC TGCCTCTCAG
118321 GTTCAAGTGT TTCTCCTGCC TCAGCCTTCA GAGTGACCT GGGATTACAA GCGCTCGCCG
118381 TCGCGCCCCG CGTGTTTTTG TATTTTTCGT AGAGACGGGA TTCGGCCATG TTGGCCAGGC
118441 TGATCCCGAA CTCCTGATTT CTGGTAATCC GCCCGCCTCA GCCTCTTAAA GTGCTTGAAT
118501 TACAGGCGTG AGTCACCGCG ACCGGCCGAA ATCGATTGGT TTTGAAGCCT TCAGTAGCAT
118561 TAAAACGAAA AGTGCTCCCA ATGCATTCCC TTTTGTCTTA AATTGGTTTC TTACAGCTAC
118621 TTTACTTGAA AAGGTGGTGG CTCGTGAAAAG AGCCTTTGCT TGGACCGTCA GAGAGACCAC
118681 AGTAATCACG CCCTCTCTCC GCGGATGCGG CCGGCGAGCT GGATGTCCTT GGGCATGATA
118741 GTGACGCGCT TGGCGTGGAT GGCGCACAGG TTAGTGTCTT CAAATAGCCC TACCAAGTAG
118801 GCCTCGCACG CCTCCTGCAG AGCCATCACA GCGGAGCTCT GGAAACGCAG GTCTGTTTTA
118861 AAGTCTGCG CAATCTGCG CACCAGGCGC TGGAAAGGTA GTTTACGAAT AAGCAGTTCA
118921 GTGGACTTCT GATAACGCGG GATCTCGCGC AGAGCCACGG TGCCCGGCCG GTAGCGGTGG
118981 GGCTTTTTCA CGCCGCCGGT GGCCGGAGCG CTTTTGCGGG CTGCCTTAGT GGCCAACGT
119041 TTGCGTGGCG CTTGCCACC AGTAGACTTC CGAGCAGTTT GCTTAGTGCG AGCCATGACG
119101 GAAAAACAGC ACAGCGGAAC ACCCAACACT AGCGCAAATA CGCCCATGAG CTGCTCTATT
119161 TATAGTGTGT AAAGTGCACT GATTGGATGA TAGAAGACGC TAAATATGAC GTTACACACT
119221 CTGATTGGTC TATCTTTAAG CCAGCAACAA TCGTGCAGTT TCACCGGCTA CTATATTCTA
119281 TTCCAACCTT ACAGATGATT ATTTAAGTGG TATTTTATTA CTACTATTAT TTTATTTTAC
119341 TTTTGCTTTG TTCCCAAGC TGGTCTTAAA CTTGGGCTCA AAAGATCTTC CCGCCTCAGC
119401 ATCCAGAGTA GCTGGGATTA CAGGGGAGCC CCACTGCGCC GGCTTGGACT TTAATTTTTT
119461 AAACCTGTCC TCTTCTACAT CTGGTTTTCA TAACCTGAAG GCTGTGTTTA TTTTCCATAA
119521 AACAAAGGCAT TGATTTCCAAA GGTATTATAA TTCCCAATT CCGTATAACC TTCAGCTCTT
119581 TAGGAAAAAA AAAAAAAGG GAATACTGCT CACCTCCTCT CCGGAAATGT
119641 ACCCTTTACG GGAATTTCTG AAACCTTTCA CAAGAATTGG ATTCTTTTGT AATGCTTTAA
119701 TTGACTTAGG AGTGTATTG AAATCTACAA AGCATCTCAA ACATAGTAGG ATTACCTAT
119761 TACTCAGAAA CATTTTCTAT GAGACGTCTT TCTCTTGATT ATGCTCTTTG AATCCTAAAC
119821 TTGCAGCGTT CTGCAGCTTT TGTTTTCTAA AGCCTAGGTG TACTCTGCCA GTCACAAAAT

```

Figure 9 (Page 37 of 74)

126/162

```

119881 GCGGTTTCTC CAGCACTGCC GCCAGGTACC ACCAGCTGGG AGTTGTTTCT CTTGCGGAGC
119941 AGGAGGTGGA CTGGGCCCAA GAGAACTGG ATAGTGGTTC GCAAGGAACA TAATTTAGCA
120001 TTGCCAAGAG CTAATGCAAT CATTTTGAAG ATCTCAAAAC ACTGAAAAGT GGATTGTGAC
120061 CTTTTTAAAT TCACAAGAGA CAGGCCACAT TCTATCTTTT GATTGGTTTA GGCTATTTTC
120121 TTGAACAGCC ATTTAGAAAG CAGATCTATC ATCCTTCATT TGCATGGAGC GTTCCCATT
120181 TATTTGAAAC CAGTTTAACC CAATAGAAAA AAGGGAGGCA GAACCCATTA TTTAAAGTGG
120241 AAACCTCTGA ATCAGATAAT TAGGAGTATT TCCTTTTCAA AAGTTGCGTT TTTTCAGATA
120301 CCTCGCTTAT TACACTAAGA AAGGTTTATA TCTTTCACAA AGGGTTTACT TACAAAAATC
120361 TTCCAATTTT GTATACCTGT GTTTCATAAC TGAAGTACCG TCAAACCAAG ATGTAGAGTT
120421 TCCAACCGTT ATTTTCCAAA TTTTGTAGAA TTACGTGAAA TATTTGAATG CATGCCCTCT
120481 CAATAAAATG GGACGTAGGA AGCACTGGTG CAGAAGATGG GTACAATACT TATCTGGGAC
120541 CACTCCATTA TTTGGTTGGC ACGTTGTTTG AAGAAAAAGG GGAAAAGCTC AGGTTACTTA
120601 GCATGGTTCG GACTTATTTG AAAACTACCA CAGCAGGAGC GGAAATAAGA CCGCATTACC
120661 TCACTCTCTG CTGTGCTGTG CTAGGGGGTT ATCCAGAATA GGATTGTAGA AGTGGATGTC
120721 GATTTAATAG TTTTTTATTC TCCCATTAGC TGAGTCTCTG ATTGGCAATG TGAGATCGTT
120781 TTAGCTTATT GATACCTTGA AATGCACCTA ACAGCCACAA ACAAGTTAA GGGTTGTTAC
120841 CATAAAATCT TATCCCCAGG GTGTGCTTGC ATTTATCACC CGTGTTTGCT TTCACACTAA
120901 GTGGACTTAA CTCCCCAGCA GAATGCCTGT CAGGGAACCG GTTTCGTGGA CCCAGCATTT
120961 AACGCCTTTC GCAGGCTTGT GAGGCCCAT AATATTTGTT GAATAAAAGA ATGAGTTGAC
121021 CATGTCATGG TCGCTGATT GCGTGTGCTG ACATGGAACA CAGGTTGTAA ACCTTAATAC
121081 CAATTTGGGG CATGTTGTAT GGATGAAAAG GGCATTGGAA ATTCTGAAG TGCATCCAC
121141 ATTGGACTGT GGAAATAAGT TGCAAGTGCA GAAACGTTTC CACACTTGCA GTTTGAGTAT
121201 TAATTGCAGC GTTTGTGAAT TCTGGTGTG TCTACGATC ATTCTTGTTC GACGTGAAAG
121261 GTATTCGCGA GACACATCGC TCTAAACAT TGCCAGAAAA TGTAATAGAG TTGATGACAA
121321 CTGGCCCTAA CACGGCCCTA AACTCGCACT TTTCTCTCCC TCCGCACTA TTCAAAACAC
121381 TGTATTTTAC ATTTCTTGCA AATTAAAAAC TAACATCTCT GGCAACGGAC CCTCTAAAAT
121441 TTCTAATAAA ACTCCTCGGA TGCTTGTGGC ACTGCATTG TAAACCGCCC CCTCTCAACC
121501 TACTCCCTAA AAAAGAGCTG CTTTTTGAGA GAGAAGCGGT ACCCTCTGAT GTTACTGGGC
121561 GGCAGTCTGC CTACAATTTT CTTCACAATG AGGCAACCAG AGCGGCTTTT TCTGTGTGTT
121621 TGCTTGCGTT GAGGGGAGCA GGACCATAGG CCCTAGAGGC CCCAGCTGC CTTCTGAGAC
121681 TGGGCGAAAC CCTCGGCAGC GCGCAGGGGG CGCTAGGGCG CGAGGGGCGG GCACTGACGG
121741 GCACCAATCA CGGCGCAGTC CCACCCTATA AATAGGCTGC GTTGGGGCCT TTTTTTCGCA
121801 TCCTGCTTCG TCAGGTTTAT ACCACTTTAT TTGGTGTGCT GTGTTAGTCA CCATGTCTGA
121861 AACAGTGCTT CCGCCCCCG CCGCTTCTGC TGCTCCTGAG AAACCTTTAG CTGGCAAGAA
121921 GGCAAGAAAA CCGTCTAAGG CTGCAGCAGC CTCCAAGAAA AAACCCGCTG GCCCTTCCGT
121981 GTCAGAGCTG ATCGTGCAGG CTGCTTCTCT CTCTAAGGAG CGTGGTGGTG TGTCTTTGGC
122041 AGCTCTTAAA AAGGCGCTGG CGGCCGAGG CTACGACGTG GAGAAGACA ACAGCCGCAT
122101 TAAGCTGGGC ATTAAGAGCC TGGTAAGCAA GGAACGTTG GTGCAGACAA AGGGTACCGG
122161 AGCCTCGGGT TCCTTCAAGC TCAACAAGAA GCGCTCCTCC GTGGAAACCA AGCCCGGCGC
122221 CTCAAAGGTG GCTACAAAAA CTAAGGCAAC GGGTGCATCT AAAAAGCTCA AAAAGGCCAC
122281 GGGGGCTAGC AAAAAGAGCG TCAAGACTCC GAAAAAGGCT AAAAAGCTG CGGCAACAG
122341 GAAATCCTCC AAGAATCCAA AAAAACCCAA AACTGTAAAG CCAAGAAAG TAGCTAAAG
122401 CCCTGTCTAA GCTAAGGCTG TAAAGCCCAA GCGGCAAG GCTAGGGTGA CGAAGCCAAA
122461 GACTGCCAAA CCAAGAAAG CGGCACCAA GAAAAAGTAA ATTCAGTTAG AAGTTTCTTC
122521 TAGTAACCCA ACGGCTCTTT TAAGAGCCAC CTACGCATTT CAGGAAAAGA GCTGTAGTAC
122581 ACAGATGAAA TCCCCAAGC AAATGCAACA CGCCCTCAAT TATATTAGAA TCACTTGGAG
122641 AGTCGATAGA ACTTTAACAT AGCCTCATCT AGTAAGAATT TACTACTCAA TCTATCAAAG
122701 ATAGCAAGGT GAATTCAAAT GCACCGAGTT AAAATCGAGT TTTAAAGTCA CCTGGGTTTC
122761 GGTAGCCGGA AGTCCCGCGT CTCACGACTC CAAGCTAATT AGTCATAACC GTATTGAACC
122821 AAGGTTGAAG CCCAGTCCCA GGCTTGAGGC TTTTATTAT ACAAGGTTAA AGTGGGGATA
122881 TTGCGTTTGG GGGTCAATAT TGCTAAAGTA GCATTTTCCG AAATTGGGTG GTCCTAAGAA
122941 ATGCTTCTGG GATAGTTGGC AAAATATATG GCTTAACCA GCCCTCTCA CAGGAGTGGC
123001 TAGCGAGCTG TCTGTCTTGG GGAAGGACGG TGACCCTGCT GCGGTGGCTG GCGCCACGCT
123061 TGCGGTCCTC TGAAAGCCCC GCCAGGTAGG CCTAGCTCGC TTGCTTCTG CAGCGCCATC

```

Figure 9 (Page 38 of 74)

SUBSTITUTE SHEET (RULE 26)

127/162

123121	ATGACAAAGC	TTTGAAACGC	AAAATGCTTT	CTTTGTGCAG	CGCCTTACCA	TGGGTGCACT
123181	TACGGGCTGT	CGACTTG GTT	TAGGCCCTTG	TCAGGACAAA	GGAGCTTAGT	TTGTTGGAGT
123241	TTTAGAGCTG	CAACCCAAAA	TCCTTTGCTC	GGTTTCTCTG	TTTTTAGAAA	CGGAAGCGCC
123301	CTGATTGGAT	ATTTGAAAA	TACTGTGCTT	AACTGGATCG	TGTTTCATCA	ATCGTGCAGG
123361	ATTTTCAACC	CTGGTGGAGC	CCACACATTC	AAAAGTGAAG	ATCCTTTTCT	CAGAACTGCC
123421	CCTTTAAGCT	TTTGCAATTT	TAATTCTGGG	GGTCAGATTT	TAATAATTGG	ACTTTTTTGT
123481	TTACATCTGA	CAAGAGTATA	TGATGAGCCA	AGTTTACTCA	CTTTTACTTA	GTGCAGTTCA
123541	ATTCTAAAAG	TTTATTTTTG	CGTGTGTGCA	TATGAGTTAA	TAATCAGTTG	TATTTTTTCAA
123601	ACGGTCTTTT	TTCAATTGTT	TTGCTTAGCT	CCTTCCATCG	TCTAAAGTCA	GGGATACAGG
123661	CACATCACAT	CCCTGTTCCC	CCTTCCTCAA	ACTAATATGT	AGCTACCTAG	GTTTATCCTT
123721	TAAAACAAAA	ATTCTCACCT	ATTTTTGTGA	GAAATATACA	TGTTTTTCTT	TGAACTAAGT
123781	ATTTTACATA	CACCTATCTA	TATACATGCA	TACTTGTGGT	TTTGTTTTTT	TAAAAAATAA
123841	AAAAAAAAAA	CACGTTATCT	TTTGAGACTG	GGTCTCAGTC	TGTGCCCCAG	ACTGGACTGC
123901	AGTGGCATAA	TCACAGCACA	CTGTAACCTC	CAACTCCTGG	GCTCAGGCTA	TCCTGCAGCC
123961	TCAGCATCCG	GAGTAGCTGG	GATTGCATGC	ACGCACCACC	AAGCCGGGCT	TTTTGTTTTT
124021	ATTTTTTGTG	GAGACAGTCA	CACCATGTTG	TCCAAGCTGG	TCTAGAAATG	GCCTCAAGTG
124081	ATCATCGACC	TCCCAAAGTG	TTGGGATTAC	GGTCACTGTG	CCTGGCCTTG	TATGCATAAT
124141	TGTTTTGTCT	TTTGATTAGG	GTTATTAATT	TAAAAACAA	AGCCTGGACG	CAGTGGCTCA
124201	CATCTGTAAT	CCCAGCACTT	TAGGAAGCCG	GATGGGCAGA	TTACTTGAGC	TCAGGAGTTC
124261	AAGACCAGCC	TGGGCAACAT	GGTGAAATCC	CATCTTGACA	AAAAATACAA	AAAATTAGCA
124321	AGGCCCAGTG	GCACGCACTT	ATAGTCCCAG	CTACTTGGGA	GGCTGGGGTG	GGAAGATGAC
124381	TGGAACCTGG	GAGGTAGAGG	CTGCAGTGAG	CAGAGATCGT	GCCACTGCAC	TCAAGCCTAG
124441	GTGACAGAAT	GAGACCCAGT	CTCAAAACAA	AAATAATAAA	AATTTTTTAC	AACGATGTTA
124501	TATACACTTC	TGCATGTTGC	TTTTCTCTTA	ACCAAACCTT	TCTAAAACCC	TGTCATGAAA
124561	AAAGAAATCC	TTCACATGGA	ATAGCATAAG	TTATTCATCC	ATTCTTATT	GATAAGCATT
124621	GATGTTTCCA	GTTACCACTG	CTGAACATGG	TGCAATTGAA	TAGAATTCCA	GGGCTGAGAT
124681	TGCTAGGTTT	TAGGTTGTAT	TTTATTATTT	TATTTATTTA	TTTATTTATT	TAGACAGAGT
124741	CTTACTCTGT	CACCCATGGT	GGAGTACAGT	GCCATGACCT	CAGTTGCAAC	CTTTGCCTCC
124801	TGAGTTCAAG	CGATTCTCAT	GCCTCCGGTC	TCCCGAGTAG	CTGGGATTAC	AGGCACCTGC
124861	CACCAGGCCT	GGCTAATTTT	TGTATTTTTA	GGAGAGATGG	GGTTTCACCA	TGTGGCCAG
124921	ACTGGTCTCA	AACTCCTGGC	CTCAAGTGAT	CTGGCCACCT	CGGCCTCCCG	AAGTGCTGGG
124981	ATTACAGGTG	TGAGCCATGG	CTCCAGACCT	GGACTTTGTC	TTCTGTTTCA	TCAGTCCTTC
125041	TGTTGGTTCA	AGCACAGTAT	CACACTGAAG	ACTGATGATT	CTATATAAAT	ATGGTAAAGA
125101	CTGTACACCC	TAAGTGTCTT	TATTTTTTAA	TTTTAAGGCA	ATTTTAGATT	CCAGCTTTCC
125161	AAAGAATTGT	GGAATGCTTA	GAGGTAGAGA	AGCCTTGGAA	GTCAATTTAGT	TTTGTTTTTG
125221	TCAGAGAAAA	TTCTGTAGAG	ACTCTGTCCT	GCTCTCACTG	AATACCATCC	CATAGTACCC
125281	CCCAACAGCT	TTAAAGGGCA	ATAATACCTT	ATGGACAGTA	TGCTTTTCTT	CAAATATATT
125341	CTAAGCCATG	GTCAATGCAA	AAGAGTGAGA	AGGAAAGTAG	AATAAGTTAT	CTAAGAATCA
125401	GTGGGTGCTC	TCTTTAAACT	GATTTATCAC	TCCCCCTTCC	AAACTCTCTT	GAAGGTCACT
125461	CTGCCTCCCT	TTCTACATAA	GAACTCCTAA	CTCCAAGGGA	GGAAGGTAAG	TTATTCTTAT
125521	TCCTTGCTTA	GAAAAAGAGA	AAATAGGTTT	GGTAAGCATC	CGCTTCTGTC	TACCATTCTC
125581	TGTGTTTCTG	TGTTTTTTAT	AGGATCATT	AATTATTGGT	TGGCTCTTGA	GAGGGAATGC
125641	AAGGTTCAAG	GACACAAGCC	TAGATCTTGC	CTGTATAGAA	CCTCATGATG	TTATGCTTCT
125701	CTAAAATGAG	GCCTGGAGGA	GACATGTTGA	AAGTGACCCA	TAAATCTGCA	GTATCTCATG
125761	TCTCTCAATG	GGGACAAGGA	GTACCATGGG	AAATAGCATT	AGGTCAATGA	CAGTAACAAC
125821	TCCCAGGTGA	GTTGATTTAT	TCTTTTATTT	ATAAAGTTGT	TAATATGCTA	CATAGTCCCT
125881	AATTTTGCCA	CAAATAGTCA	TTATTTTAA	TTTATATTTT	ACTATTGATA	AATGAAGGAA
125941	AAAATGAGTA	GCAGTTAAGC	AGTCCATAAA	CCTACATATA	AAGCAAATTG	GAGATTTTAA
126001	AATTGATTCT	GGATGCTTAA	AATCCTTCTC	ATTGAAAAAA	AATTTCTGAT	TAGAAGATTT
126061	CAACATTCTT	TAAACTGAGA	AGCATAACAT	ATAAACAGAA	AACCACAGCA	AAACAAAAAT
126121	GCAAAGCTCA	ATAAATGAAC	ACAAAGTGAA	CACCATAATA	ATTGCCACAC	AAGTAAAAAA
126181	ACAGAAAATC	AGCCAACCTT	CCCAGAGCTG	CCTGATGCTT	GCTTCCAGTC	ACATTATCAC
126241	TCCATCTGCC	CTAAACATAA	CCCCTATTTT	GATTTCCAAT	GCTGTAATTT	AGTATGCCTG
126301	TTTTTGAAAC	ATATAAAATG	GAAATAAAAC	AAATGTAATC	CTATGTACCT	GACATATTTT

Figure 9 (Page 39 of 74)

128/162

126361 ACTCCAGAAC ATTAGGTTTG AATAGATTCA TCTGTGTTGC TGTGTATAAC TTTAATTCAT
126421 TTTTATTGTT ATGTAATATT CCATGTTATG AGTGCAACAA TTTAGGTGTC TACTGTTGAT
126481 GCATATTGTC TTCCCTTTT CAGCTAATAT AAACAATACC GTGAATATTC CTGTGTATGT
126541 GTCCTGGTAT ATATAGGAAT ACATATTTTG TTTGTATACC TAGGAGAGGA ATTGTTGGGT
126601 CAAATGCTAA ACTCTTTTG AAAGTGGTGA TATTAGGTTT ACATGCGATG AAATGAAAAT
126661 TAAACCACA GTTATAAACA GCATGGATGA ACCTCACAAA CCTAATGTTG ATGGAATCTA
126721 GCTGGGAATT CCTGTTCTTC CATATACTTC CCAATATTTT TTTCCAATTA AAATTGTTAA
126781 TCTTTGAAG ATGTTATCCA TTGTGGCAGA TGTGCAGTAT TATCTCATT TGGTTTTATT
126841 TTACATCTTT TGCCCATTTT TTCTTAATTG GATTGTATAT CAGTCGACTT GGGCTGCCAT
126901 AACAAAAATA CTAGACTAGG TAGCTTGAAC AAAAGGAATT TATTACCTCA CAGTTCTAAA
126961 GGCCAGGCCA GAAATCCTAA ATTGAGGTGC CAAGAGATTC AGTTTCTAGT GAGGGCTCTC
127021 TTATTGACCT GAAGATAGTT GCTGTCTTAG ATTGTTTGGT GCTGAACAGA ATACCAGAGA
127081 CCAATAAATT TATAAGAAT ACAGATTTAT TTCTTACAAT TCTGGTGGCT ATAAAGCCTA
127141 TGGTCGAGGG GCCCACCCTT GGCAAGGGCC TTCTTACTGT TATGGCAGAT GTGAGATGTC
127201 ATCTCATATT CAAACCACAG CAGTCGCCTT TTGTGTCTTC ATGTGGCCTC TTCATATGCC
127261 CATAAAATGA CCTCATGTCT CTTCCTTTTC TTATAAGGAC ACCAGATCTA TCAGACTACT
127321 GGCTTACTCT TATGACCTCA TTTAACCCTA AATATCTCCA TAAAGTCCCA AAATCCCCTAT
127381 CTCCAAATAT AGGCACATTG GGTGTTAGAG TTTCAACATC AATTTTGGGG GAACACAATT
127441 TAGGCCAAAA AGATTGTGTT TTTTCTTGTT GGTTAAGAT AGCTGTCTTT TTGTCCCTTT
127501 TGTCCTTTCT TTTTTTTGA GGTGGACTCT TGCTGTGTCA CCCGGGTTGG AGTGCAGTGG
127561 CGTGTCTCA GCTCACTGCA ACCTCCACCT CCTGGGTTCA AGAAATTCTC CTCCTCCCAA
127621 GTAGCTGGGA CTACAGGTGC ATACCACCGC GCCCTGCTAA TTTTGTATT TTTGATAGAG
127681 ACGGGGTTTC ACCATGTTGG CCAGGCTGGT CTCAAACTCC TGACCTCAGG TGATCCACCT
127741 GCCTCGGCCT CCCAAATGC TGAGATTACA GGTGTGAGCC ACCAAACCTG GCCTGTCTTT
127801 TCTGTTTTAA GTTTTTAAAT TTTGCTCAGC AACCTTTTAT CCATTTTATG TGTTCAGGTT
127861 ATTCCTCTG TAACTGTCT TCACTCTGTC AGAGGCTGGA GTGCAGTGGC ACAATCACAG
127921 CTCCTGTCAG CCTCCACCTC CCAGGATCAA GCGATCCTCC CATCTTATCC TCCTTAGTAG
127981 GTGGGACTAC ATGTGCAGGC CACCATGCCC AGCTAATCTT TGTATTTTTT TGTAGAGATG
128041 GTGCTGTTGC CCAAGTTGGT CTCAAACTCC TGAGCTCAAG CAATCCATCA ACCTTGCCCT
128101 CCCAAAGTGT TGGGACTAGA GGTGTGAGCC ACCACTGCAC CCAGCCAATG ATATCTCATG
128161 ATGCATTAAA GTCATTAAAT TAGTGACTC AAATTAAGCA CACTGCCCTT TTATGCACAA
128221 CCTTTTTTGT ATCTTATTTA AAAATCATT TTCTATTTCA AGGTCATGAA GATCTTATTT
128281 TATAATACCT TCTGTGAAA TTAGTTCTCA AGACTACCCT CACTTCTAAC ACCAATTATA
128341 AGTTGGGAGG TCTGTGGTTC CCAATCAACC TTAGGTTAGT AATTTGCTAA AAGGACTCAC
128401 AGAACTTGCT GAAGCTGTTA GCCTCATGTT TACAATTTAT TATAGGATAT ATAGCTTATT
128461 ATGTCATTCC AATGCAATGT AAAATTATAC AACTACTTTT AAAAAGATTT TAGCATTTGA
128521 CCCAACAAAT TCACTCTGAG GTATACAAAC AGCAGATATG TGTGCACATA TATACCAAGA
128581 CACATACACA GCAAAATTCA TTGTTTGTA TAGTTGAAA GGGGAAACAA CTCAGGAAT
128641 AAAGATTAAA ATCAGCTGAG AAAAGAAACA CACAAGGCAG TATTATGGAT CGAATTGTAT
128701 GCAGATCTCC CTGCCCCCA GAAGATATGT TTAAAGTCCC AACTCCCAGT ACCTCAGAAT
128761 TGTGGCCTTA TTTGGAATA GGATAGTTGC AGATATAATT AGTTAAGATG AGGTTATAGT
128821 ACAGTATGAT GGGCTGGTGA CTTAGAAGAA GTAGTATATA TATATTTTTT AATAGAACTA
128881 GTATTCTTCT AAGGTGGTCA CGTGAAGACA GACACACACA GGCAGAGACT GCGGTTATGC
128941 AGCTGCAGGT CAAGGAATGT CAAAGGTTGC CAGCAAGTAC GAGAAGCTAG GAAGAGTCAA
129001 GGAAGGATTT TCCTACAGGC TTCAGTGGAA GCATAGATCT AATGATACCT TCATGTCAGA
129061 TTTCTAGCTT CCAGAACTAC AAGAGAATAT ATTTGTTGTT TTAAGCCACC CTAGCTTCTA
129121 GCTCTTTGTT ACAGCAGCCC TAGGAACTA ATATAGGCAC AATCCAGGCA AGTTCCAAAT
129181 ATGAGCTTCC AGTTGTCCTC TCCAGTAAT ATGAACAGTA TTACTTTCCC AGCATTAAATG
129241 TGTGACAATA CACATGACGT ACAGAGCAGT CCCCCTTAT GCACAAAACA TATGTTCCAG
129301 GACCTCCAGT GGATGTCTGA AACCATGGAT AGTACTGAAC TCTATATAGC TGTTTTTTCC
129361 TATACAGACA CAGCTATGAT AAGGCTTAAT TTATAAATTA GGCACAGTAA GAGATTAATA
129421 ACAATAAATT AGAATAATTG TTAAGTAAT ACTGTATAAA AGTTAGGTGA ATGTTTATTT
129481 CTGAAATTTA CCGTTTATTA TTTTGGACT GCAGTAGACC ACAGGAECTA AAACCATGTA
129541 GAAACCGTAT ACAAGAGAAC TGTATTTTAC CCGAGCCTCA GTGTGCAGTT TTAATGGCCT

Figure 9 (Page 40 of 74)

SUBSTITUTE SHEET (RULE 26)

129/162

```

129601 GCCATGGTTG ACTGCTCACA TGGCCGATCT TTTAGTCTAC CTCCACAGGT AGAGCTGATA
129661 CTGTGTGGCT CAAAGTTCCT ATTATAAATC ACATTGTTGA CTGTGTGGTG GTCAAAACCT
129721 CCAGGTAAAC AAAGACACAC TTATCAGTGA GAACATTTCA AGGGTCTAAA ATTCATCTCC
129781 CAGTAGCTGA GGGCAAAGGC TAGACCTCTT TTTGGGTAAG ATAAATTTTT TACCATATAC
129841 TTTATTTTGC TTTTCATGTT TAACCTTTATT TTGCTTTTCA TGTTAGTTCC CCTGGAATTG
129901 TTTTGTGTGT ATAGTGTGAA GTAGGGGGTC AAGTTTCTTT TTTTTCCTT TTTGTTCTTT
129961 TTCTGTTTTAA AAGGCTATAC AATTGTCCCA TGCCATTTAT TTACAAGAGT CCTTTCACCA
130021 TTGTTGTATG GTGCCACTTT AGATGTAAAT CAATGTCCAT ATTTGTTTGA CCCTGTTCCA
130081 TTCGTTTGTG TATTTTGGGA CAACACTGCC CTGATTATTG TCATTTTATC AGTTTTGATA
130141 TTTAATAAAG CAACAGATTT GTTTATTTTG GGCCCTTGGA TTTGTGTATT AAATTTGAAC
130201 CCTGTTTGTG AATTTCTATA ATAAAGCTTA TTGGGAATCT GATTAGGATT ACAATGGTTT
130261 TGTAGATCAG TTTGGGGACA ATTAATACCT TTAAAATATT GACCGCTTCA ACTGTAAATA
130321 TACTCCTCCA TTATTTAGTT TTCCTGTTTA ATTTATCTGA GTAATACATT ATAGTTTTCT
130381 TCGTAGAAGT CAGATACGTA GAAAATTCOA AGCCCAAGTG CAATAGCTCA TGTCTGTAAT
130441 ACCAGCACTT TGGGAGGCCG ATGTGGGTGG ATCACCTGAG GTCAGGAGTT TGAGACCAGA
130501 CTGGCCAACA TGGTGAAACC TCATCTCTAG TAAAAATACA AAAATTAGCT GGGTGTGGTG
130561 GCGGGCACCT GTAATCCCA GCTAATCAGGA GACTGAGGCA GGAGAATCGC AGTGACCCAG
130621 GAGGCAGAGG TTGCAGTGAG CCAAGTTCCT GTCACTGCAC CCCACCCTGG GCGACAGAGC
130681 GAGACTTCGT CTCAAAAAAA CAAAAAAAAG AACATTCAA TAATCAATGT AGATAATTCA
130741 AATAACTAAA AAATGAACAG TTATTAATAAT ATCAGGATAT AAAAGCAAAA AAATCAATAA
130801 CCTCCATATA TACAAAATGG CCAGTTAGAG AAAAAAATAA GAATAGGCGA GACTTAAAAA
130861 GGCTGGGAAT CTCCCTGAAA ATCTTTGAGA GCCTTGGCCC TGCCCTCAGG GATTTCTCTG
130921 GCTTCATGCC CAGATATGGG TACAGTTCCT TGTTTAAAAA AATTTTGCTC CATCAATCAA
130981 CAAGGGGCTC CTTCTCAGA GCACAAGGAC CTCCATAACA CCGGACACTA GATGTCTAAG
131041 GGACACCTCT TAAGGAAGTT AGACTTCCAA AGAATGGTGT TTCCTCTGTC CCCAACTCT
131101 GGAACTCACA GCACAAGTGC TCCTTGGAGT TCGGTTTCAA ATCTACAAGG CTGTCATGGA
131161 GGTTGCAGAC CAAGTCCGTG GCCTCAGTGT CCGGATGTAC GGTGGCCTTG GCACCTGAAT
131221 GTGAGAACAT GACCTCCCTG AAACCACCAC AAGTATTGTT TCATGTTATG TATGTTTTTT
131281 CTTATCTGAA ATTCTTTTTC TTTAAAAATT CAAATTACAT ATTTTTCAG CCCCTGAACA
131341 AGCTTCATGA GCATTTATTG AACCACACGC TTTTAAACC TACTGAACAC TTTGCTCTAT
131401 GTTGTCATTC ACTATCCACC AATTATTTAA TTATTGATCA ATATTGTTTC CTTAGTGTTG
131461 GGATCATTTA TGCATGTATT TCTTTTATAT TGCATATTTT ATATTTCTGC ATTACAGTTA
131521 TTACATATTA CTTTGTCTAC AGTAATAGTT CAGAAGTGTA CATCCAAAT TTAGCTGTGA
131581 AGTGGATGGA CTGAGGCAGA ACTGGAGGCA AGAAAATGTC ACAGTAATC TAAAAAGAT
131641 GATGTACAAT TAGAGCAAGA GAGTAGCAGT GAAATTGAAG AAAAAATAGT GCGTTTGAGA
131701 GAAAATTAGG AGGTAGAATC AACAGATTAG ATGTAGGGAT GAGAAGGGTC AAAGATGACA
131761 CTAGGGTTTT TAACTGGAGC AAGTAGGTAG ACAGAACATT TCTTCTGAA AGGGCAGGTC
131821 AGATCATGTG TTGTCTCAA GGCATGAAG AGTAGAAAGC CTGGGACAGA TCCTGAGATG
131881 ACCAATACCC ATGGTGCAGG GAGAGGGAGG GAGATCTGCT AAAAAGACTG CAAATGTCAG
131941 GATAGTAGAA AATCATGAGT GTGTGATGTC CTGGAAGTTG AGACAGTATC ACATTTGAGA
132001 ACATTTAAAT TGGTAACCTC GACAAAACCT GGAGGCCAAC TGTGAATGCC CATGAGAGTG
132061 AGAAGCTCCC ACACCTTTGT GGGCATCAGA AAGCCCACCA GGTTCTGCA GTGAAGATCT
132121 GAGAAGGATC CTCTTGTGGC TTTGGCAGGG AGAGAAGAAT TATTATGAAA TACACCCCAG
132181 AACCTTCTTC AAAACAAAGG CCTACTCTCA AGGGGAAAAC ATTTTGCCAG AGTCTTATCC
132241 CAGCTGGGAG AAGGTAATTC TTCCCACTGC AGCCTCATCT AGGCTTTCTG TCTCACTTAA
132301 GGGAAAGAAA TTAGTCAACA GGGATCAGAG CTTATGAAA ATAAATTGGA AATGGTGCAG
132361 CCAGGAAAGG AGCAAAGGTC TGAGGAGGAG GAGAAGGAGG AAGAGGAGTT GTATCATTAT
132421 AAATACTTGA GGAAGAGGAG GAGAAGGAGG AGGAGGAGGA GTTGTATCAT TATAAACACT
132481 TGAGGAAGAG GAGGAGGAGA AGGAGGAGGA GGAGTTGTAT CATTATAAAC ACTTGAGGAA
132541 GAGGAGGAGG AGAAGGAGGA GGAGGAGGAG TTGTATCATT ATAAACACTT GTGACGGTCC
132601 CAGCCCCAAG ATATAGGCAT GCTAATAAAC TGAGGCTTAA CACTTTGACT ACAGAATGCT
132661 GCTTCTCCCT AACACCATCA AGGCTCCAAC TGAATAACAA TGAATTATGA ATGAAAGAGC
132721 TGTAAGGAGA GACAAAAGTT AGAATGAGAC AAGTATTGTT ATCTAGAGAT GCCAAGAAGG
132781 CAAGGAAGAT AACTAAAAAG GCACTCTGGA TTTAGAAATA GGAAGTCATT AGTGACCTTG

```

Figure 9 (Page 41 of 74)

SUBSTITUTE SHEET (RULE 26)

130/162

132841 TAAATAATGG AGCCAGAGGA ATACCAAGGG CAGAAGCCTC ACTATAGTGT GTTGACCTG
132901 TCAGAGGTCA GGAGGTGTAA CTGACTCTCC CACAGTGTGG CTTTGGAAAGA GAGAAGTCAG
132961 CAGCTGCATG GAGATTGGG AGAGGGAAAG CTTTTTTTTT TTTTTTTTAA TTGGAAAAGA
133021 CTGAGCTATG TGTAAATAGA ATAAGACAGG AAGAGTGTAG ACACAGGAAA GAGGGCAGAC
133081 AAAACAAGT GCACAGTTAT CTAAGGGAAA CAATGGGATC AAGCTGCAAG TATATAAACT
133141 TGTCTTGATA GAAGAATCCT TGATCTGGTT TATTCAGTGT TTGGTCCAAA CCCACATCCC
133201 TGTTCTGCCT GTCTCTGACT TGCTCTGTGC CCCAGAAGCC CAGCTTCTAC AGATAGCATT
133261 AGCTGGGCAG CCTGCCCCC TTGCAACAGC TGGATTTGGC CAGTGATCAG CCCAGCAGGA
133321 ATGTAGATGG CAAAGGAGAG AGAGGTTAGT GTACTTATTC CCTGCATCAC CCCCCTGCTT
133381 GGTGGGCAGC TCTTCCTCCA CAGTCCCAGC TCTGGCCTAG CTCTGGTTAC AGGTTCCCTC
133441 CCATTGCCTC TTCAGATTTA AAGGTGTGTC TGTCAAGGTA TAACTGGGAG CTAGAAATTG
133501 CACTGAAATT GAACAAAGAA TTTTATGGGA ATGGTTGTTA ACTAGTTATA AGAGGACTGA
133561 AAATGGAAAA GTGGAACAAA CGTATCAGAG ATAGTAATGA CAGAAAGCAA CTACCACCTC
133621 CAGGTTTAGG AGAACAAGGA AAAGATTCTT TGAAGAGATC CCCAGAACTG GGACCTCTGA
133681 GGAGTGTATG CTGGACCACT GATGATGATA TGCTGTAGA TAGAGGCATG ATGAGGCTGA
133741 TTTTAGGAGC ATGGAAGATC TCCAACTGA AGCCAACTGC TGTACTGGA TTCAACTGCC
133801 ACTGCCAGGT TGAAGAACCC ATTCTGTGAG GATGTCAACA AACAAAGTGG GAAATCTTTT
133861 CACATCCTTC CAGCCCTCTA GTCTTCCTCC AGTGCTTTCT ATTGGTAGGG TTTGGGGAGG
133921 TGGCTAGCAA AGCGGTATTG GAAAAGATAG AAGAGACTAA ATCTTCATAA CCAGCACAGG
133981 GTGACACTGG ATCACTACTG TTGCTGATCT TGGGCTGCCT CATATCCCCT GTTCTTCCCA
134041 TTAGCCCTGT CACAACTTTG TAGATATCCC TTCATTATAT GCCCTTCATA TATTCTTTTG
134101 GTTTAACTTT TTCTGTTGGA ATCCTAATAT GGCACCTCCT CATTTTTCAG GACCAAAAAGA
134161 GTATAAAAGA TTATCTTTTA CCAAAAAAAA GACAAAAAAC TGATCTAATT CCTGATTGTA
134221 TCATTACACA ATCTATACAT GTATCAAAAT ATCACATAGT ACCCCATAAA TATATACAAC
134281 TGTGTCCATT AAAAATAAAA ATTAAGAAAA AGATGGTAAA TATAGCTCTG TCAGGCAGTG
134341 GAGGTTTTAC CACGATGGCT GTTATTTCCC CCATGAAGGG GGGAGTGAGG GAGCAGCTGA
134401 AAGTAGGTGC TTATAGGGGT ATAGAGGGGC TCAAAGCTTT GAGAGAGGAG AATGTCTGAA
134461 AGAGCTGCCA AATAGCATGC AGGTCCCATG GGGGCAGAGC CTCTGCTCAT TCACCAGTGC
134521 CTCTTCAATA TCTACACTTA AGCCTAACAC AAAGTGTGTG CTTAATAAGT ATTTGCTGAG
134581 TATGTAAAGT GGAAACAGAA CCAATCTGGC AAACCTTGTA GGACTGGTGG GCAATGAAGA
134641 TCAGTCAGGT AAAATCTGTG GATATAAATT TATATTGATC AAAAAATTCA AGGTTAGGTG
134701 TTTTCTTCA GTCATGCTCA ACGATGCTTC AGCCATGCTC AACTCTTCTG TAGCCACAGA
134761 AAAAAGTTTA CCCATAATCG AGCTGTGTCT GTGTCTGAAT AATGAAAAGA CCATGATGCA
134821 AGGGAGTTGG AGACACAGAA ACAGTGTGTT AAGTAATGGG TAATGGAAGC ATGCTACCAG
134881 GGAAAGGAAA GAAGTGGCAA TAGGAAGGAA CAGAGATCTG TGGTCTATG TCCCCTGAGC
134941 ATATTCACAT GTTAAAGCTA ATTCAGTTTT CAATCATCAT TAAAATTTTG TTCCTAAATA
135001 TATGGCCATT ATTTTCCACA ACCACACTAA AACTTTATTA CCTCTGGCAA GTGACTATGC
135061 AAGTAACTAA GAGCAAAAAT ATCCCAACT ACCATTTGAG CTATCAATTT AGGGAAAGTC
135121 ATCTGGCTAT AATCTAAGTG ACCCTCCACT GAATGTCAGT ATCTTTGCAT ATGTGATTTA
135181 AATCTGGGCC TTCGCAACAC CATGAACTGT TCTGTCTTG AATATCCAGA TTGAAGGAAA
135241 TAATCTGAGT AGTTACGAGT CCTGAAGCTA GAAAGATGGA AACCCTATT GCTCATCAGA
135301 AAGCCTTAGA GCTTGGGCGC TGGCGGGTCC TGTCTCACCG GGACAGAGGG GCTCTTCTCT
135361 CCCCATCTGA TAGTCTGATA ACTAGAGAAG CCGGCCAACT TATTCTCAA GAAGGAGCCA
135421 TCTTAGTTCC TCCGAAATG TTCATATTTA GAAATTATTG TTTGTGAGTA ATTTAACCCC
135481 TTAATGGGCT TGCCTTGTGG TCCATACCAC TGAGTGCAGA GCTTGCTGG AAGAATTGTG
135541 AGGGCCATT CATCTCCAG GCAGTAGAGT TCAGTACTTC TTTAAATTG CTGCTGAACT
135601 CTGTATTGTA AAAGAAAGAA TCATTGGGT GTGGTAGCTC ACACCTGTAA TCCTAGCGCT
135661 TTGGGAGGCT GAGGTGGGAG GATCATTTGA TGCCAGGAGG ACCACTTGAG ACCACCCTGG
135721 GTAACATAGC AAGACCCTGT CTTTAGAAAA AAAAAATACA ATAAAAATAA TACAATAAAA
135781 ATAAAAGCAA AAAGAAAGAG TCCATCTTAG GGACAGACTG TAACTACTCA CTGGAGCTTA
135841 CCTTTACATA GTTCAGGATC AATTATAATA AAACACTTTT GTGCAGATTC AATAGGATTA
135901 TTTTAATCCC CATCATCTCT CTGAGTTTCC AGTCAGTTTC TCTGCATGTA GACACCTTTC
135961 TCCAGCCCAC CATTGTCTCT CCTCCTATAG CTCCACCAAC AAATCAGAAC TTTTCTAAC
136021 TGCACCTAGT GCACCTAGAG TCTACTCCAG AATGCTCATG GAGAAAGTTT CTGAAAGGTA

Figure 9 (Page 42 of 74)

SUBSTITUTE SHEET (RULE 26)

131/162

136081 AAACCTCTGAA TGATATTTGT AGCTAAAGGG AGACTTGCTA GAGACAATAA GCTAATAGTT
 136141 GTAGACTTCA GTAGAAGAGG AATGACACTG CAATGTCAGG GTGCAGGACT TCAAGAGGGC
 136201 AGAGTATGGA AACCCAATGG GAAAAATGCT CACCAGGAAC ATGAAGAGAA GGAATTACGT
 136261 GTAAGGATTT CTCAATGTGT TCCCAAATTT GCCCAGCAGA GGGAGGCCCTC GGGTTGATGG
 136321 CAGGCTGACC ACACAATTAA AGAAGGCTGA ACCTGGGGGC TTTTAACAAC CATCGTGGGC
 136381 TCTACTGTAA GCATTTAGAA AAAGAAAGTT ATCCATTCAA AAATATATAT ATTTTAAAC
 136441 TTCAGAACAA AATTATGAAG AGCTATATTT ACTTTTCTAC ATTCTAATTT TTATAAATCT
 136501 GAGTATATTT TGCATATATT GTTATAGTAC ATATTCAATT TTGTATTTTG CTGTTTTTAC
 136561 TTAACCATTT TTAGTAGATT ACTCTGTGTT CATAATAATC ACTTTTTTAA AACTTTTTATT
 136621 TTTATTTATT TATTTTTTTT TTGAGTCAGA GTCACACTCT GTCGCCCAGG CTGGAGTGCA
 136681 GTGGCGTGAT CTTGGCTTAC TGCAACTTCC ACCTCCTGGA TTCAAGCAGT TCTCCTGCCT
 136741 TAGCCTCCTG AGCAGCTGGG ATTACAGGTG TGCACCACCA AGCCCGGCTA ATTTTGTAT
 136801 TTTTAGTAAA GACGGGGTTT CACCATGTTG GTCAGGCTGG TCTCCAATC CTGACCTCAT
 136861 GATCTGCCCA CTTGGCCTC CCAAAGTGCT GGGATAATCA CTTTTTATGC TGCATAATTC
 136921 TTCAGATTTG TCAGTACGAC TGTATTTACA CTCATTGTG TTATTAGAAA GAATTCCAGA
 136981 ATATTTTGGC TGCCCTAATT AATTTTACAA TTAATATGAT TTTGAAATTG GGTATTGGCT
 137041 CCTTCTGAAT TGGTTTATTA AAATATATTC TAATGTAATT TATGACATTT TCATCATATT
 137101 AGCATATTTA TTCTGTTAGA ATTTTCATAA TTATAAAGCT ACAAACGTGA TGTGATATAG
 137161 CTTGTAACCT TATCTCATAA CTTTATGCAG TTACAAGTAG AAATAAATG TTCCCTCAA
 137221 GATTGCTTAA AATTTTTATTA TAAAGAACTG TAAAAACAA AATCACTAAA ACACCTCCCT
 137281 TTTTTCCCC CAAAATGCAT GTTTCCATTT TAACAGAACC CGTATTTAAT CAGCAGATTT
 137341 CTATGGTGGC TAGATTTGTA GACTAAATAT TAAAAGTCCC AAAGCAAATG CATTTTTCTC
 137401 TTAAATTTTA CTGACTTTTT TTTTTTTCT TTTTCTGAGA CGGAGTCTTG CTCTGTCGCC
 137461 CAGGCTGGAA TGCACTGGCA CAATCTCGGC TCACTGCAAC CTCCGCCTCC CGGATTCACG
 137521 CCATTCTCCT GCCTCAACCT CCCGAGTAGC TGGGACCACA GCGCGCCGCC ACCACGCCCA
 137581 GCTAATTTTT TGTATTTTTA GTAGAGACAG GGTTCACCG TGTTAGCCGG GATGGTCTCG
 137641 ATCTCTGAC CTCATGATCT GCCCACCCTA GCCTCCCAA GTGCTAGGAT CACAGGCATG
 137701 AGCCACCGCG CCCCCTTAC TGACTTTTAT CCAAAGAAA TATAAGAGCT CTTTCATCATA
 137761 ACGTATGTTT CTTGCTCTTG TTATTAAATA TGACACATTT AGACTTAAAC TGATTTGAAG
 137821 GTTTATGACA TTGTTTAAAG TATTACATAA TTAATTCATA AAGATAATGA CTAGTTTGAA
 137881 CTACTGACAG CTCACACATC ATCAGTTGAA CAGCAGAAAG CTTATTAAGC TACTTTCTTA
 137941 TGTTTCTGTC TCCCAGCTAC TAAAAGAAAC GAAACCCTTC CAGGTGTTAA GGCAAACTT
 138001 TCCTCCCCCT TTCTTCTATA AATCTGATTC CATGTTAGTG AAATTTCTAC TGATGGCTTT
 138061 GGTTCCTCT ATAGTAGAAT AGAGATCCTA TGGCAAAAGT CATGTCTGAC ATGGTAGCAA
 138121 ATAGAAATGG GGAAAAGGAA GGTCTGCAAG AGCCAATGTG GGAAATGGGG AGAGGACTGA
 138181 CTACAAAAAC CCAGCAGGAA TTCCAGAAGA AAACCTCTCA GGACGGGCAC ATTGGCTCAT
 138241 GCCTGTAATC CCAGTACTTT GGGAGGCCGA GGTGGGCAGA TCACTTGAGT CCAGGAGTTT
 138301 GAGACCAGCC TGGTCAACAT GGCAGAACCT CATCTTACA AAAAAATAAA AAATTTGTCA
 138361 GGCCTGGTGG CATGCACCTG TAGTCCCAGC TACTCAAGAG ACTTAAGTGG GAGAATCACT
 138421 CGAGCCTTGG AGGTGGAGGT TGGTGAAGCC AGATCACGCC ACTGCATTCC AGCCTGGGCG
 138481 ACAAAGTGAG ACGCCATCTC AATCAATCAG TCTCTCGAA AAGCAACATT ATGGAGAGAC
 138541 AGGATTCCGT CAAGGCCTGG GGCACACAGG AAAATATTAA GGCAGAAGAG AGTTTCCTCC
 138601 CCACACCACA CCGTATCCCA CAGGCACTGC GGATGTGCAT ATGCAAGAGG GGTGATCCT
 138661 AAGAATTTAG AGTCACAGAG GAGGAGGCAC CAAGCAGACT GTGGAGAAAG TCATGACCAG
 138721 AAAGGGACAG AATGTAAAGC TTCAGCTGAT TATCTGGCCT CAGGGATTCC AGAGGAACCTG
 138781 GTCCCAATGG TCTCCTGGTG ATGTAGGTTT TTAGGTTTCT TTTACAGGGG TTTTCTGGGA
 138841 GATCGTTGAC CCAGTTAGCA TTCAAGCAAC TTCCACCCTG CACTTTTATT CTTTCCCTT
 138901 CACCTGCTTA GGTTTTATCT GTCCAGGCAA TAATAATAAA ATTATTGAGC CCTGGACATG
 138961 TACCTGTAAA GCTCCTTAAA GATGATGCCT TCTAATCCT CATTCACAG ATACAAAAAC
 139021 ATTACAATAA AATGACTCAT GCAAGACACC CAGGTAGTTT ATAGCAGCTA ATAAAAACAG
 139081 AATAACTATA AAATATGGTA AGTTTATAAA AGTTACATTG AGTATACCTT ATAAGAACTG
 139141 CTTATTGAGT TTGCCTAATA ACCACACAGC ACAATAATAA TATGTATATA TTTTAAATA
 139201 TGTGTAAATA TGTGTAACAC AAACCTGTAG AAGGTATATC TGAGTACAAC CCTATTCTGT
 139261 TTGGTTACCT TTTCTAGTTC ATTATGTAAG TGGCATAGCT ACCTAAGGAC TTATGCTTAT

Figure 9 (Page 43 of 74)

SUBSTITUTE SHEET (RULE 26)

132/162

139321	AAATGTTACT	CAAAAAAATA	CAGAGGACAT	ATGTGGATAG	ATAATGGAAG	AGATAAGATA
139381	GGTAGGTTGA	AGGGTTGGGC	TGCCCCCTCA	CACCTGTGGG	TGTTTCTCGT	TAGGTGGAAT
139441	GAGAGACTTG	GAAAAGAAAG	AGACACAGAG	ACAAAGTATA	GAGAAAGAAA	AAAAGGGGTC
139501	CAGGGGACCG	GTGTTTCAGCA	TACGGAGGAT	CCCACCGGCC	TCTGAGTTCC	CTTAGTATTT
139561	ATTGATCATT	ATTGGGTGTT	TCTCGGAGAG	GGGGATGTGG	CAGGGTCAAA	GGATAATAGT
139621	GGAGAGAAGG	TCAGCAGGTA	AACACGTGAA	CAAAGGTCTC	TGCATCATAA	ACAAGGTAAA
139681	GAATTAAGTG	CTGTGCTTTA	GATATGCATA	CACATAAACA	TCTCAATGAC	TTGAAGAGCA
139741	GTATTGCTGC	CAGCATGTCC	CACCTCCAGC	CCTAAGGCAG	TTTTCCCCCTA	TCTCAGTAGA
139801	TGGAATATAC	AATCGGGTTT	TACACTGAGA	CATTCCATTG	CCCAGGGACG	AGCAGGAGAC
139861	AGATGCCTTC	CTCTTGTCTC	AACCTGCAAAG	AGGCGTTCCT	TCCTCTTTTAA	CTAATCCTCC
139921	TCAGCACAGA	CCCTTTACGG	GTGTCGGGCT	GGGGGACGGT	CAGGTCTTTC	CCTTCCCACG
139981	AGGCCACATT	TCAGACTATC	ACATGGGGAG	AAACCTTGGA	CAATACCTGG	CTTTCCTAGG
140041	CAGAGGTCCC	TGTGGCCCTC	CTCAGTGTTC	TGTGTCCCTG	AGTACTTGAG	ATTAGGGAGT
140101	GGAGATGACT	CTTAACGAGC	ATGCTGCCTT	CAAGCATTTT	TTTAAACAAAG	CACATCTTGC
140161	ACAGCCCTTA	ATCCATTTAA	CCCTGAGTTG	ACACAGCATA	TGTCTCAGGG	AGCAGAGGGT
140221	TGGGGCTAGG	GTTAGATTAA	CAGCATCTCA	AGGCAGAAGA	ATTTTTCTTA	GTACAGAACAA
140281	AAATGGAGTC	TCCTATGTCT	ACTTCTTTCT	ACACAGACAC	AGTAACAATG	TGATCTCTCT
140341	CTCTTTTCCC	CACAGGAGGT	GATGGCCGGA	AGAACATGGC	AGAGGGCAAA	ACAAAACAGC
140401	ATTGGGAACA	AGCTCTGTTT	AAAAGGAGAC	TTGTGAACAG	CAAAGAGTAG	AAAGGGTTCT
140461	CTTACAACCTG	AAGCCCATGG	AAGACAAATG	TGTACTGCGT	GAGTTTTAAG	GCAATAGGAG
140521	TAGTGGGACC	TAGGGCACAC	CAGAGAGCAT	ATTAACCTCT	AAACTTTTTAA	AAACATTATA
140581	TCTGTGGAC	ACAGTGGCTC	ACACCTTAAT	CCTACAACCT	TGGGAGGCCG	AGGCGGGCGG
140641	GTGTAGCTTG	AGCCCAGGAG	TTGAGAGCCA	ACCTGGGCAA	CATGGCAAAA	TCCCCTCCCT
140701	ACAAAACAAA	CAAAACAAAA	ACAAAATTAG	CCAGGCACGG	TGATGCGTAC	CTGTGGTCCC
140761	AGCTACTCAG	AGGCTGAGGT	GAGAGGATCG	CTTGAGCCCC	GGGAGGTTAA	GGCTGTCAGT
140821	AGCCATGATA	ATGCCACTGC	ATCTCAGCCT	GGGCAACAGA	GGGAGAACCT	GTCTCAAAAC
140881	AAAAACAAA	ACACACCATA	CCCAACCACA	ATGCATCTGT	CTTAAGTACC	AGTACCACAC
140941	CCCTCTACTC	ACTACTAAAT	AGGTGAGTTC	CCAATCCCTG	GTAGCAGGTT	TAAGCATGTT
141001	ATATTAAAGG	TCTTAGGCTA	GTGACTCATT	CACTCATTAA	ACAAATACTT	ATTGTGCATC
141061	TACTATAAAC	TAAGTACTGT	GCTAGGTACA	AAAGCAAATA	ATCTAAGCTC	TATAAACTTT
141121	ACTTTCTTCA	TCAACAAAAT	GGAGATGTTT	TAGGCATCTA	CTCATCATTC	TGAGCTCCAT
141181	CTTTTGTGAC	TGTAGTTGGC	AGAGCTTTTT	ATCAGTTTCT	CTAAATAGCT	CTACCAGTCC
141241	CTGGTGGATG	CTGGCATGCC	CAAAGGATCC	ATCCTGATGG	CCCTGTCTGC	TTACCTTACC
141301	TGCCTGCCTT	TGCAGCACCG	TCTGCTCTT	CTGCAGGACT	TCCCTTATCC	TTTGGGGTCT
141361	TGCTGCTCTT	AGGCTGCTCT	GCTTGTTTTG	ATCTGCTTTG	CATCACATGT	ATGTAAAGGT
141421	CCTTTCCTTA	TTTACCCATG	ACCAAGGTAT	TATGAGATTC	TGGAATTTCC	CCAAACCACA
141481	TTGATTGCTG	GGAGAATAGA	AGAAGTGGAT	TACAAGTGA	ACTTAGAAGG	GGAGTATTCG
141541	AGAAGACGTC	TCTGCAAATC	CATTTAGAGA	GACCTTTCTC	CAGTGGTGAC	TCAAAGATGC
141601	AGCTCCTTTC	ATCCTGTGGC	TTGGCCATCT	TCAGCACATG	GCTCCCAAGG	ATGTCCTCAG
141661	GATGGTCTCT	AATCCAAGGA	GCCTGAAGAG	AAAAAAAGGC	ATGGAGTATT	GTGAGTGGTA
141721	GGTGGTTATG	GACCAGTTAT	GGAAGAATAC	ACATCACTTT	TGCCCACCTT	CTACTAACCA
141781	GAACCTCACAC	AGCCATAGAC	ACTGACAAGT	AGGACTTAAC	AAGAATCTAA	TTTTGAGTCT
141841	AGGAATACGA	CTGTAGCAAA	TATTTAACAG	CTTCAAACAC	AGGTGCATTG	CTATCACTAT
141901	GCTTGGCCCA	GGCCTGTCTC	CCTTTCCTGC	CATGTCACAG	GGGCCAGCAT	TTATGTCTAG
141961	ATTGGGTTGG	TTGGGATATT	AAGACAATAA	TGAACCAATA	CAACATCTTG	AGCATAAAAC
142021	CAACTGATAC	AATGATGTAC	AAGTCAGATG	ATTCTGATGA	TTATGAATTA	TGTCAATAAA
142081	AGAAATGTGA	TAACTAAGGT	AATTTTTGTT	TTGGCAAATT	TTTGTTTGTT	CATGACAGGA
142141	TGAAATCCTG	TCATTTGTAG	CAACATGGAT	GGAATTGCAG	GATACTACAT	TAAGTGAAT
142201	AAGCCAGAAA	CAGAAAGTTA	AACACCACAT	GTTCTCACTT	ATATGCAGAA	GCTAGCTAAC
142261	TAAGTAAATA	AGTTTATCTC	ATTGAAGTAA	AAAGTACAA	AGAGATTACT	AGAGGCTGGG
142321	AATGGTAGGG	GAAAGAGATG	ATAAAGAGAG	ATTCATTAAA	ATAAGTTACA	GCTAGATAAG
142381	AGCAATCAGT	TCTAGTGTTT	TATTTGTACT	ACAGAATGGC	AATAGTTAAC	AGTAATAAAT
142441	AATTTCAAAG	AGCTAGAAAA	GAGGACATTG	AATGTTTCCA	ACACAAAGAA	ATGAGAAATG
142501	CTTGAAATAA	TGGATATTCT	AATTAATTAC	CCTGATCTGA	TCACTATACA	CAGTATGTAT

Figure 9 (Page 44 of 74)

SUBSTITUTE SHEET (RULE 26)

133/162

```

142561 AAAAATAACA CTATGGGCTG GGCAGCTGG CTCACACCTG TAATCCCAGC ACTTTGGGAG
142621 GCCAAGGTAA GCAGATCACT TGAGGTCAGG AGTTAGAGAC CAGTCTGGCC AACATAGTGA
142681 AACTCCATCC CTACTAAAAA TACAAAAATC AGCCAGGCGT GGTGGCATGT GCCTGTAATC
142741 CCAGCTACTC AGGAGGCTGA GGCAAGAGAA TTGCTTGAAC CCAGGAGGCG GAGGTTGCAG
142801 TGAGCCGAAA TCGCGCCACT GCACTCCAGC CTGGGTAACA GAGCAAGGCT CTGTTTCAAA
142861 AATAAATAAA TACATAAATA AATATTTTTT AAAAAAGAA CATCACTATG CACCCCATAT
142921 ATACATATAA TTATTATGTC AATTGAAAC ATAATTTTGA AAAATGAAAA AATGAAACAC
142981 AAATATGAAT CAATCCTCTC CAAGTTGATA TACTTAAAAG GAAAAAGTC CGAGGGCTTA
143041 AACTATTCAA TCAAAATTTT ATTAATATGC TATAGTAATC TGGAAAGTAT TTCAGAATGA
143101 ATTGGTATAA GGTTAGACAC AAAGATCAGT GAAACAAAT AGAGAACCCA GAAATAGATT
143161 CACACATCTA TGGACAACTG GTTTTGACAA AGGTGTCAAG GCTATTTAAT AAGTAAAAAA
143221 ATCGTCTTTT CAGTAAATGT TTCTTGAACA AGTAGACATC CGGTGTGGGG GAGAGGAGCA
143281 GGAGCCTTAC CTCAACTTT ATGCAAAAAA TAACTCAAAA TAGACCATAG ACTTAAATGT
143341 AAAAGCTAAA ATTATAAAAC TTCTTTAAAA AATAGGAGAA AATCATCAAC ACCCTAGGAT
143401 TAGCAAGAT TTCTTTAAAA CAAAACAACA GGTTTATAGT TTATAAACA TAAATAACAA
143461 AATGATAAAT TTCATCAAAA GTGAAAATTT GCTTTTCAAA AAACATTATA AAATGAAAAG
143521 CAGGAGGCTG AGGCATGAGA ATCACTGGAA CCCGGGAGCT ACAGGTTGCA GTGAGCCAAG
143581 ATGGTGCCAC TGCACTCCAG CCTGGGTGAC AAAGTGAGAC TCTTCCTAAA AAATAAATAA
143641 ATAAATAAAT AAATAGAAAA GAAAAAGAAA AATCACAGGC TGAGAGAAAA TATTTATAAT
143701 ACATGTATCT GACAAAGGAC TCGCACTGG AAAATATAAG GAACCTTATA ACTTAGTAAAG
143761 ATGACAAGCC AAAACAAAGA GTAAAAGTTT TCAACAGACA TTTCACAAAA GAAACATAC
143821 AAATGGCCAG TATGCACATG AAAAGATTTT AAACATCATT AGTTACTAGG GAAATGCAAG
143881 TCAAAACCAC AATGAGATAC TTCACATTCA ACAGAATAGC TAATGTTAAA AGGACTGACA
143941 ATCCCAAGGG TGAGCAAGGG ACTACTCTCA TATATTGTGA ATGTAGAGG
144001 CATTTTATGA TATAACTGAA TTCAGTTTTA TGTATAACTG AATTACGGAT ATGAGAACTCT
144061 CAAATGAGGA CGAATGGTTT TTACGCACAA AACATGAGAC ACAAATCTGT AAGAAATATA
144121 AAGTCGTGAC CACGTCCTTT CAGAACTTTA ACCTGTTTGC TGAAGTACGT CAGTAACAAT
144181 GGCAGGGAAA GGGTATCTTA AATTTACCA CAGCCTCAAA GAGGCCATTT CGTGGATCCG
144241 CTGAGGCTTG GAGTCGGCCT TCTGACCACG AGTCTGCGG CTATGAAAGA GGAAGCCCGG
144301 GTTCAGGGCG TCCTCGCGAG TCGCGCAGCC CGCCTGCTC CAGCTGGGGA CACAGGTGGT
144361 CACGGCGCTT TCCAGCTGCA GATCCAGGCG GCAGCCCAAG ATTTGGTCCA GCCGCCAAGG
144421 GGTGGCTCGA GTGACTGACG GGCCTTGAAC GCTCCAGGA CCCACATCTG GAGAGGGAGG
144481 TGGGGGTGGG GTGCTGAAGT CATCTTGGG GCCCTGGGG GCGGGCATGG ACCTGGGTAA
144541 GGCCAGAGAA ATTGACACCT CGTGACATCC CTGGAAGAGA AGTACGTTCA GTGTCACTCC
144601 AGAGCTGAAA GATACCGCCT TCTGGCTGGT CCCTCCTCAC CTACATACTT TTCTAATTTG
144661 TCTGGAGCAG GCCGGGCATC TGTATTATCT GGTATTTTAA ATATCTGGTT ATTTAAAAGC
144721 TCTCCATTAA ATTACATAC ACGAAAATAA AAATTA AAAA TTTTAAA AAAAAGAAAC
144781 AAAAGCTCTC TAATGACCAA GTCCTACACG ATAGTGAATA AATTTTTTTG TGTGGTCCCT
144841 AAAATTGAGT TCATGCCTTT TCTGAAGTAA TAGACGCCCA GAGAAGGGAT CGACTTACCC
144901 ATCATGCCAC AGAGATTAAT TGGCCCCAGA ATTCTTTAGC AGACCGTGTA TATGAACGTC
144961 CTTTGCAATC ATATAAATTA ACTGGGAAAA CCTCATTTAG TATGTTACAT GCCTAGCGTT
145121 TTGTGCCTGA ACACCTTACA AGAACCAGGG ACTATTGCC CAATATTATA TTTCAGGAAA
145181 GGAAGGCCCA GACAAATGGT GTCACCTGGT CACTTTCACC CAGTTGGTAA ATGAAACCAG
145241 AAATTATAGC TGTACCACAG AAAGGTGAAA ACGTTTCTTT TATAATTTCA CATACAATCT
145301 TTAATGGACC CAGTGTCCTA CACATTAAAG CAAGTGCTCA GGAGTGACAT CAAGATGTAA
145361 AAAATAGTCC TGTCTCAGG GAGTTTAGGT CTTGGAGAAA AGAGACCCAA GGAGACACAA
145421 GACAAAGGGG AAAGAGAAGG AGCCTGAAG ACTGAGGACC CTGCCTGTGG ACTGAAGTGA
145481 GGATGGGGAC ACCCGATGCC CGGAATATGA CAGTTTGGAG GGGCCTGAAG GACTCTTCTA
145541 TTCTCTATCA GAAAAACAGA ATTACTCTCC TAACCAGAAA AGGTATTTCA ATTTATATTT
145601 TCCATCACAG CACTTTTCTG GTGATAATTT AATGTGTTTT AAAAAATGTA TCACAGTGAT
145661 GGCTTGGTGT GAAATAAATA ATAAAATTTT AAGAATTAAA AAATATAAAA ATCTTTTATA
145721 TAGACATTAG GAGTTACAAG GATAACTGTG AATTATAATT AGTAATTAAA TTGAAATACT
145781 GATTATTTTC ATTTTATTTT AATTATTTAA TAAAACCTAT TTAACATTTA ATATTTATCA
145841 GTAATTAAAT CTAATTGTTA ATATTTATTA TTATAAATTA TTTTAGAATT AAAAAAAGT

```

Figure 9 (Page 45 of 74)

SUBSTITUTE SHEET (RULE 26)

134/162

145901	GTAGAAGCGA	GGCATGGTGG	CTCAAGCCTG	TAATCCCAAC	ACTTTGGGAG	GCTAAGGTGG
145961	GAGGATTGCT	TGAGCCCAGT	AGTTCAAGAC	CAGCCTGGGC	AACATGGAGA	AACCCGTGCT
146021	CAATACAAAA	AAATGAGCCA	TGTGTGGTGG	TGCGTGCCTG	TATTCCCAGC	CATTCTGGAG
146081	GCTGAGGTGG	GAGGATGACT	TGAGCCTAGG	CAGTCAAGGC	TGCAGTGAGC	CCTGATCTTG
146141	CCACTGCACT	CCAGTCTGGG	CAACAGAGCA	AGACCCTGTG	TCAATATACA	TATGGACAAA
146201	CTTAAATTTT	AAAATGAAAG	CATACTACTG	ATACAGAATT	GAGTAGAGAT	GCAAAAGCTAG
146261	TCCTATAACC	AGAACAATAA	AGATAAAAAG	GAGAGTGGAA	GAAGGTATGT	CATGAATTTT
146321	ATGATAAATG	GCAATTGCAA	ATATCCTGTA	GCAGAACAAA	ACAACAAAAT	TGTAGATAAA
146381	ACATATCCAA	CCCTTTGGAA	GGCCAAGGAG	GGAGGATTGT	TTGAGCCCAG	AAGTTGGAGA
146441	CCAGCCTGGG	CAACATAGTG	AGACCCTGTA	TCTAAAAAGG	AAGAAAGAAA	AAAAAAAGAA
146501	AGGATGATAA	AGTAGACAAT	ATTGAAAGCC	ATTTTCTGCA	AATACATAGT	GAATTTGATC
146561	AGTAATTTTC	TTCCAACAGT	GCAAAAATGA	ATAGATATTA	GTTGCCTGAA	ATAAAAATCA
146621	AATATCCAAC	AAAAAATATT	GACTATCTAA	TAGTATCTAA	GCTAGTAAAT	TTGGCCAGTT
146681	ATAAAATGTC	TTAAATTTTT	ATTTAAAAAA	AGAAAACCAT	ATTTATAAGA	AGAGGTGATA
146741	AAGAGAAATT	ATTTCAAGTTA	TGAAGATTTT	GTTAGAAAAC	TATGAGAAAA	AACTATTTTT
146801	TTGTTTTCAA	AAAGTGAAAG	ATTAAGTTAC	CAAACAGTTG	CTAAAGAATA	CCAGATGGCT
146861	GAGCGTGGTG	ACTTATGCCT	GTAATCCCAG	TACTTTGGAA	GGCCAAGGCA	GGAGGATCAT
146921	TTTAGGCCGT	GAGTTCGAGA	CCAGCCTGGG	CAGTGTAGCA	AGACCCGTCT	CTATTAAAAA
146981	AAAAAAGAAA	AAAAAAGAAA	AATACCAGAC	CTTGCTAACA	ATAGCAAGAA	TCAATTAATT
147041	CAAAATTTGA	AAAACGTAA	TTTATTTAGC	TTTAGAGTAC	TCTCGTGATA	TGAGATTGCC
147101	AAATTAATAC	TTTGGGTGCA	TTTCTTTTCT	CAAAGGACTT	GCAAATTTAC	AAAGAAGTGT
147161	TGAAGAAAAG	CCACACATTG	GCAGGTAATG	TTTGCAAAAG	ACAGATCTGA	TGAAGAACAA
147221	TATTTTTTAGA	ATATACAAAG	AATACTTAAA	ACTCAACAGT	AAGAAAATAA	CCTGATTTAA
147281	AGCAGGCCAA	TGACCTGAAC	ATCTGTTTAC	CAAAGAAGAT	ACACAGATGC	AAGTATGCAT
147341	ATGAAAAGAT	GCTTGACATC	ATGTCATTAG	GGAACTGCAA	ATTAAAACAA	GTAGATACCA
147401	CTGCATACCT	AGTAGAATGA	CCAAAATTTA	GAACACTGTC	AGCACCAAAG	GTTGCAAAGA
147461	TATGTAGCAA	TAGTAACCTG	TTCATTACTG	GTGAGAATGC	AAAATGTGCA	ATCACTTTGG
147521	AAGACAGTTT	GGTGGTTTCT	TACAAAAGTA	ACCATACTTT	TACCATAAGA	TTCACCAATC
147581	ACACTCCTTA	GTATTTATCC	AAAGGAATTG	AAAACCTTATC	TCCACACAAA	AACCTGCACA
147641	TAGATGTTTA	TAGCAGCTTT	ATTCATAATT	TATCCAAAAC	TTGGAAACAA	GATGTCTTTC
147701	AGTAGGTAAG	TGGATAACTG	TGGTACTTCT	GAATAATGGA	ATGTTATTTA	GAGTTAAAAA
147761	GAAATGCATT	CACTTTGGGA	GGCCGAAGTG	GGTGGATTGC	TTGAGGCCAG	GAGTTTGAGA
147821	CCAGCCTGGT	CAACATGGGA	AAACCCCAAT	TAGCCGGGCA	TAGTGGCGTG	AGCCTGTAAT
147881	CCCAGCTACT	CGGGAGGCTG	AGATATGAGA	ATCGTTTGAA	CCTGGGAGAT	GGAGGTTGCA
147941	GTGAGCCAGT	GCCACTGCAC	TTCAGCCTGG	GCAACAGAGC	AAGACTCCTC	TGTCTCAAAA
148001	AAAAAAGAAA	AAAAAAGAAA	AAAAAAGAAA	AGAAAAGAAA	AAAGAAAAG	AAAAAGAAAA
148061	GAAACGATCA	AGCCATGAAA	ACACATGAAG	GAAACTTAAA	TGTATGTTAC	TAAAAAGCCA
148121	ACCTGAAAAG	ACTGCATACT	ATATGACTCC	AACTGATGCA	GGGCAAGCAA	GCCAAAAATT
148181	AGGGCTTAGC	CCGGGAAGAA	TTCAAGGGTG	AAGTGGTGGT	GTTAGCAACT	TTTACTGAAG
148241	CAGCAGTGTA	CAACAGCAGA	ACAGGTACTG	CTCCTTGCTG	AGCAGGGCTA	ACCCATAAGT
148301	AATGTGCCCA	GAGTAGCAGC	TCAGGGGCAG	TTCTGCAGTA	ATATACCTGC	TTTTAGTTAA
148361	GTGCATGTTA	AGGGGGATTA	TGCAGAAATT	TCTAGAAAAA	GAGTGGTAAC	TTCCGGAGTAG
148421	GTACAGAGGA	AAGAAGTCGA	TAATGTCCTG	TTGTTGCCAT	GGCAACGAAA	AACTGACATG
148481	GCGCTGGTGG	GCGTGTCTTA	TGGAGAGGTG	CTTTAACCTC	GTCCCTGTTT	CGGCTAGTCT
148541	TCAATCTGGT	CCGGAGTAAA	GTCCCTGCCT	CCGGAGTTCA	CTCCTGCTTC	CTGCTTCACA
148601	ACTGTATGAC	ACTCTAGAAA	AGACAGTAAC	TATGGACACA	GTCAAAAGAT	TAGTTGATAG
148661	AAATTGGGTG	ACAGGAAGTG	TTGAAAAGGC	AGAACACAGG	ATTTTTAGGG	CAGTGAAACT
148721	TCTGTGATAC	TATAATGGTG	AATACATGAC	ATTATACATT	TGTCAAAACC	CATAGAAAGC
148781	ACAACACCAA	GAATAAACCC	TAATGTAAAT	TACAGACTTT	CGTTGATAAT	GACGTGTCAA
148841	TGTAAGTTCA	ATTGTAATAA	ATGTACTACT	GTGGTGCTGG	ATGTCTATGG	TGGGGGGACA
148901	TTTTTGCTTC	AATAGTTACA	GTTGAAGTAA	ATGTTTGTGT	TTCCACCAAT	GCATATGTAG
148961	AAACTCTCAC	ATTCAATGTG	ATGGTCTTTG	GAGGTGGGCT	CTTTGGGTGA	TAGTTAGGTT
149021	TAGTTGAGAT	CCTAGCAGAT	CGAGTCTTCA	TGATGGGCAT	GATGGGACTG	GTCCCTTATA
149081	AGAAAAGACC	AGAAAAGCTAG	CTCTCTCTTT	GCCATGTGAA	GACATAGCAG	GAAGGTAGCC

Figure 9 (Page 46 of 74)

SUBSTITUTE SHEET (RULE 26)

135/162

149141	ATCTGCAAGC	TAGGAAAGGG	CCTTCACAAA	GAATCAACTC	AGACCTCAGA	ACAGTGAGAG
149201	ATAAATTGTC	GTTGTTTAAG	TCACTCAGGC	TGTGGTATTT	TGTTTCAGCA	GCCCAACCTA
149261	AGACTGTTAA	TTGGATTAGA	AATTTCTTTT	TGGGGATGGT	GTGTGGCGGG	GGGTGCGGGG
149321	AGTACCTTTG	TTAAGCTTTT	ATATCAATGA	GTTTGTAGGC	TTTTCTTTTT	TGGTCATGTA
149381	CTAGGACAGT	TTAAATAGTA	TGAGTGTGAA	GGAGATTGTT	GGTCATCTAT	TCGATGTCCC
149441	TTCTCTGTTT	TTTAATATGA	GAACTCCTGA	TTTTCAGCCA	ACTACCCTGG	AAAAAAGCT
149501	AATCTTTCTG	ACTTCTTAAG	TGTGGCCATG	TACTAAATTC	TGGCTAATGC	AAGGCAAGCC
149561	AAAGGTTTTA	TGATAGGTTT	TAGGACACTA	GAGTAAAAGA	GAGCTGTTGC	ACACATGCTC
149621	TTCAACCCTAC	TTTTGTGTCC	TTTTTTCCAT	CCTACAACCTT	GGGTTGTGAG	TATGATGGCT
149681	GGAACTTTAG	TGGCTCTCTT	GGATCCCAGG	GGTAATTGAG	GGGTGGCTGG	AAGGAATCTG
149741	TGATTTTCTG	GAGTTTCCAT	ACACAAACAA	GACCTGGATT	TTCTGGGCTT	CCCAGACTTC
149801	CACATCTAGA	CTTGCTTTAA	ATGGGAGAGA	AATAAACTTG	TTTCAGCCAC	TGTCATTTTG
149861	GGCTATTTTA	TAGAACTTAA	TCTAATCTTC	AAGGGTACAT	GAATTGCTTT	TCCTTAAAAA
149921	AAAAATCAGC	CATAAAATCA	TCTTCTTTTT	TCTTTTGTTT	CCCACATTAT	TTAGTTGGAG
149981	CTCTGTAAC	TTTTTTTTTT	TTTTTTTTGA	GACAAGGTCT	TGCTCTGTCA	CTTAGGCTGG
150041	AATTCAGTGG	CATGACCATG	GCTCACTGCA	GCCTTGCCCT	CCTAGGCTCA	AGCAATCCTC
150101	GTCTCAGCCT	CCTGAGTAGC	TGAAACTAAG	GCACATGCCA	CCATGCCCAG	CTAATTTCTT
150161	TTCTTTTAGA	GATGGGAGCC	TTGCCCAGGC	TAGTCTCAAA	CTCCTAGCCT	CAAGTGATCC
150221	TCCCATCTCA	GCCTCCCAAA	GTGACAGGAT	TACAGGTGTG	AGCCACCATG	CCTGGCTGCT
150281	CTGTAAGTGT	CTGAATTTCA	TTTTGTATTT	ATCAGTCTGT	TTAGATTTTC	TTTCCCTTCT
150341	TGGGTCAGTT	AGGCCATTGG	TTTCTTTTTA	AAGGTTTTCA	AATTTATTTG	CATCTAATTC
150401	TTCAAATTAC	TCTCAAAATT	ATTCCAGTAT	ATATTCTTTT	GTTCTTATTT	TCTTCTGTAT
150461	TCTTTATTAA	AATAGCTAAT	GATTTATCTA	GCAGGACTTA	TATTCTTTCC	ATAACTTTCC
150521	TGCACCCCAA	TTAATCTCCA	ATTTTATATT	TCTTCTGGCC	TTCCCTATAG	TTTCCACAGG
150581	TTTATTTTAT	TCATTTTTTA	AACTTTTAT	TTAATTGTTT	ATTTTATTAT	CATTCTTTCT
150641	TATTCAGCAA	TCTAAGTGCT	TAGGGATATA	GAATTTCCCT	TAAGCAGCAT	ATGCTAGGCT
150701	TTAACAATGT	TAGGGAGGCC	TCCCTTTTCT	GGGGAAGACC	ACACTTACAT	TAACACAGGA
150761	CTGTGGGATG	CCAAGAGGTA	GAGAAGAGCT	TATGAATATC	CAGATTACAT	CTTCACTGAT
150821	CCTGCACAAA	GGTGGGGTTC	CTCGGTTACC	CACTGGGTCC	TATTACCCAA	GTCTGGGTCA
150881	GCATACCGAG	ACTACGGGTA	TATAGAACAA	GTGCAACTGG	CGATAATCCT	TCTGTTGGGG
150941	AGAAAAATCT	TTTTTTTCTA	TTCATCTTAG	GTCTCCATC	TGTGGCCCTA	TCAAGTAGAC
151001	TAACAAAAGA	CAGATTGACA	AGACAGAAAC	AAAGCATGTG	CATTGTACAA	ACACAGGGGA
151061	GTA CTGAGAT	GAATACTCAA	AAGAGGATTT	AGAACTTGGG	CTTATATAGC	ATTTTAAAGAA
151121	AAGAATACAT	TTTTTAAGTG	ACAAGGAAGA	CGAAAAGGAC	TTTGAGTTTC	TAGTGCAGTA
151181	AATTGTGGGA	AGGCAACTTT	TTCTTTCCCT	TTTTTTTTTT	TTTTTTTTTA	AAAAAAGAC
151241	TTCTCTGGTG	CTATGTCCAG	GCTGATAAGA	GTCTAAAGTC	TCTGGTGACT	AACTTTGTGT
151301	CTTCCCGGAG	TAAGAAGACA	CCTTCACAAT	TTCATATCCT	GCTTTTAGGC	AAACAGGGAG
151361	AGGGCAGAGG	TGTTTGTGTT	TTTTTAATCT	ATTTTTTTTC	TCAATTGTCT	TCAACTCAAA
151421	ATACTTCTTA	TGCCAAAGAT	GGCATATTCT	GCTACCCTTC	ACTTACTACT	TACAACCCAG
151481	CCTCTATCAT	CATAATTAGA	ACTTCTGACC	CTGGGGAACA	TGGGCAATAG	TTTGAATCTT
151541	TTTATATCTC	CCTTAGGCAG	AGATGGAGGC	CCAGCCATGC	CTCTGACATC	TAGACACAAC
151601	TGTTGCTTCA	TTTCTCCTAT	TCTCAGAGGT	GATGTTGTAG	GACTTCAACA	AATATCAGTA
151661	AACATTAATT	TTTTTTTCC	TTGAGGCACA	GCATGATCTT	GGCTTACTGC	AGCTGCTGCA
151721	GGCTCAAGCA	ATTCTCCTGC	CTTGGCCTCA	CGAGTAGCTG	GGTTACAGGC	CCCTACCACC
151781	ATGCCCGGCT	AATTTTGTGA	TTTTTAGTAG	AGACAGGGTT	TCACCATGTT	GGCCAGGCTG
151841	GTGTTGAACT	CCTGACCTCA	AGTGATCCAC	CTGCCTCAGC	CTCACATAGT	TCTGGGATTA
151901	CAGGCGTGAG	CCACCATGCC	TGGCCATCAA	TTTTTATGTC	AACTCTAAAT	TATAACATTT
151961	AGCAATTTTG	TGACTTTTTA	TGGTCATCAT	TAATGTTGTT	TATGTTTTAG	TTGTAGTCCCT
152021	GTCATTACTC	ACTCGGGTAT	GGTAATTTGG	TCTTTTTCAA	AATGAAGTTA	AGGCTATTTT
152081	GCTCTTCTCT	GAATCATAAT	AAGAACTGCC	AACAGCCATT	TCAGCAATAA	CTATTTACTG
152141	AGATTTTAAA	ATATTTCAAG	GTAATTTGGT	CTAGCAGACT	GGAAAATACC	AAATCTTTT
152201	CCAGAAGTGA	ATCCCCATC	AAAGTTCAAT	TTTACTCATA	ATTCCCTTTT	CATTTGAAGC
152261	ATCTCATTGT	AAGCCAGTCT	TAACCTTTCT	CTCACACTTT	GCTTGGCTGT	TTCTCAGGTA
152321	GAACTCAGTA	AGTCTGGTAG	CCTCCAGGAC	TGCCGCTTAG	ATTATTAAAC	AACATGTCAG

Figure 9 (Page 47 of 74)

SUBSTITUTE SHEET (RULE 26)

136/162

152381 TGGTTGGAAG AGTCAATGTT ATTTTGATTT TTCTGTTTTG TTTTGTTTTA AATGCAGTTG
152441 GCGGATAATT GCAGCTTTCT TTCATTCCCT ACATGAGTTC AAATGGCAGC AAACAAACTA
152501 GGAGAACGCA GACCTTCTGA CTTGTGGGTA CCCCTACTCA TCACCTGAAG ACCCTTGGAA
152561 ATCAAAGCCC TGACCCATTA AAGACGGATG GAGACAGCAA CATACGATCA TCACTATTAT
152621 CTTGCTTTGC CCCAGTCCAG GTTAACCATC TGTGGTATTT TTAGTTGCTA AGTCCATATA
152681 TTCAACATAA ATCAATTATA TATCCACTAA AATCTCAGCA CTAGTCTAAC TACTAAGGAA
152741 ATGACAGCGA AGAAAACAGA CCAAACGTCT GCCCTTATGG GATTATATTT ATTTTCTCTG
152801 TGCTGGTTAA ACCAAGGAGC TTCTGCTCTT TTCCTTAGTC ACCTGGGGGA GGCAGAAACA
152861 AAGGAGAATA TTGATAAACC TGGAAATAGG GCCGGAGAGT ATCAGAGAAG GAAGCCTTCG
152921 GGAAAGTAAA GATGTGGCAG CCAGTATTCC CGTTATAAAA GGATACAACCT CCGGCCTCAT
152981 AGTCCAGAAA AATTCCCACA AGCAGGGGCT GCTCATGCAG ATGAAGGGAA GTTGGGGGAG
153041 AAGTAAAGTC TACATAGCCT TTCTTTTTGC ACAGCCTGAG GGTCCAGAAT CCAGACTGAG
153101 GCTCTTGCTT CATGCCAGTG CCCCTCTGCA CATTTTCCAT ACAAACTCCT AAATCCCATC
153161 CGGTTCCCTC GCCAACATCC ACTTCAAAGT AACGCTCTCC TGAGGTGAAG CCTTCACAAC
153221 CCAAGACACA GGGGAAGGCA GTAAATCTCC TGGAAGATGT GTCCTGATTC TCCTGGGTGT
153281 ATCCACGAGT CACTTGTCTC CGATCCTCAG AGAGAATTAG TTCGTGATGA GCTGTATCTG
153341 GATCCAGAGT CACACTAACT GCAAAACAAA ACAAAACAAA CAAAATAAT TTTGTGCTG
153401 TGAAGAACAC AGGTTATTTT ATTTTATTTT ATTTTGAGAT GGAGTGTTGC TGTCAACCAG
153461 GCTGGAGTGC ACTGGCACTA TCTCAACTCA CTGCAACCTC CACCTCCTGG ATTCAGGCAA
153521 TTCTCCTGCC TCAGCCTCCG GAGTAACTGC GACTACAGGT GCGCACCACC ACAAGTGGCT
153581 AATTTTITTA AATTTTCTGT AGAGATGGGG TTTCGCCATG TTGGCCAGGC TGGTCTCAAA
153641 CTCCTGACCT GAAGTGTTCC ACCCACCTCG GCCTCCCAA GTGCTGGATT ACACAGGTGT
153701 GAGCCACCAT GCCCAGCCAC AAGTTATTTT CAATAAAACC AGCCTGTGTT CAAACCCAAC
153761 TATTGTTTCT TATAAACTGG GTGAGCTTAG GCAAATCATT TAACTTTCTG AGCCTCAGTT
153821 TGTTAACTAT AAAGTGGAAA TTACCGTATT TGTTCAGAG AATGGTGGGT AGGATTGAAT
153881 AAGCTTATGT TTGCTTAATG CTTGGTAAAA TTCCTGGTAC ATGGTAACCA CTAATAAGT
153941 GGTAGTTGTT GGGGTGATCA GGCCCAACAC CAGGCCGTGG GGGCTACAAA GTCCGGCGGG
154001 GTCAAAGGAA TGAGAAAAGA CAAGTTAAGA GTGCATAAAG TGGGTCCAGG GTGCCAGCAC
154061 TAGATTGGAG GCTGCAAAGG CCCTAAGCTC TGGGAGCCCA CACTATTTAT TGGTGATCAA
154121 ACAAAGAAGC AGGTGGTGAG GACGTGAGGG TAAACAGGTG AGGGCATGAG GACATGGGGG
154181 TAGAAAGGTA GTGGTGCAAT AAGCGTAGCT GTGACAGTTT AGCATTITCT TTGACACATG
154241 TAGAATATAC TCTGCTGCTT GAGATAGTAG AGGACACGTT TATGAGTGAA AAGCAAGGAA
154301 CCAACAAGTC TGTGCACTTT CCAGAGGCTA TGAGGGGTTT TATGCCCTGA GCCCTGGGTT
154361 CCATCCAAGC CACAAGGGGT TTTATGCCCT AGGCTTAGAT TTGTGGTGCG GCAGGGCAGC
154421 CTTCCACCAT TTGGCACAGA GCTTGGTGTG CCAAAGGCCA CGAGGGGTTT TGGACCTTGG
154481 ACCCCGGACA TCTTCCAAGA CTCTTTTACA TTATGACAGA CAAGCCAGTC CTGCTTCAGC
154541 TCTTCTAACA ACATGTAGTA ATAATGATAT CATCAACATC ATCTTCGTCT TAATTATTCA
154601 AGGATGCCAA GGTACAGAAC TAACCTGTTA ATATGGTTAC CATCCTGTCC AAAGTTCTTC
154661 TCCCATGCAG GACTTCCAGG AATCATGAGA CAGTTGAGCA GAAAGATACC TTTTCCCTTC
154721 TCTACTGAAT AACCACCAAC ATTGAGAATC AGAGAGGGAA AATGACTCAG CTAATGTCTT
154781 AGCTTGTTAT TGGAAGACCC AGGTCTCATG ACACATGCCT AGTCCCATGA CTTTTAATTG
154841 TAAGCTCTTC TCTTCCCCTC CAGATAATGT TCCATAAGCA TTAGTATGAG ATAATAATAC
154901 ACTGAGGACC AATATACATG AAAAAATACA GACTAGAATC AAACAAGACA GAAAAAGAT
154961 CTGATAACCT AAAGTGAGAT ACTGAACAGT ATGCAGTTT AAAAATAAAA AATGGTAATA
155021 GGATGTTCTA ACAAGAGAGT TAAGAAACCA CTGTGCTACT GAGTTAAATG TTGATCAGTT
155081 GGTCTGTGAC AATTAAGGAA TTCAAGTATT CAGAAACACT TCCTGTGCTG GATGCTCTCT
155141 GTTTGTCTCT CCAAATAATC CCTCACTTTT CCCTGTCTTG CTCTGTGCCC AGGAAGGCTG
155201 ACATGGACAG ATTAACCAGG CTTTCCGCC TCTGGCTTGG TTCAGCCAAT GGAAGCACC
155261 AGAGGAGACC ATAGGGCACA AAGAAGCAGC CTTGGGAGTA TTCAGTACCC CAGTCCCACG
155321 CTATGATTTG GAGGGTCTGC ATTCCTCTGC CTCTGGGCAC ACTCTAGTAT AGTTACAGCT
155381 CCTACACCT GCCACTGAG GCCCAGAGGA GGTGATGGCT CTCTAACTGT TCCTAGTTCT
155441 GGGTGCTTCC GTTCTCTGT GGAATTTCCA ACTCCTCACC TTTGTAAATA CCCTCTTTT
155501 TCAAACCTTA TTCAGTTAGC TTTTATCAGC CTGACTCACA GAAGTTGGG GTTTCAATTC
155561 ATATTACCTG AATGACCCAG GAAAACCCAT GTTGAGAAAT TAAATGTTT ACGGGGTGGT

Figure 9 (Page 48 of 74)

SUBSTITUTE SHEET (RULE 26)

137/162

155621	AATACCACTT	AAGAGAAAAA	ATATCAATTG	GATTTTTTAA	ATTCCACCTA	TCTATTGGTG
155681	TGACACATCA	ACAAAAACAT	ATAGAAAGAT	TGGAAGCTAA	AAGATAGATA	ATATAGTCAT
155741	ATACTGTTAT	AGTATTATAT	CAAAAAGATAT	TAAGTCAGAG	CATTATTAAG	AATGGAAGAA
155801	GGGCCAGGTG	TGGTGGCTCA	TGCCTGTAAT	CCCAGCACTT	TGGGAGGCCA	AGGCAGGCGG
155861	ATCACTTGAA	GCCAGGAGTT	CAAGACCAGC	CTGCCCAACA	TGGCAAAACC	CTGGCTCTAC
155921	CAAAAATACA	ACAATTAGCT	GGGCATTGTG	GCACATGCCT	GTAATCCCAG	CTACTTGGGA
155981	GGCTGAAGCA	CAAGAATCAC	TTGAACCGGG	GAGGCAGAGG	TTGCAGTGAG	CTGAGATTTT
156041	GCCACTACAC	TACAGCCTGG	GTGACAGAGA	GAGATTCTGT	CTCAAAAAAA	AAAAAAAAGA
156101	AAGAATGAAA	GGAGTCACCT	AAAAAGATA	ACACAATTTT	AAACATAAAT	GTACTACATT
156161	ATTAGTGAAT	TCATGTTTAG	AATGTGTGTA	ATATACAAAG	CAAAAATTGT	AGAATTATAG
156221	GAGAAATGGA	CAAATCTACA	ATCATCATGG	GATGTTTTAA	CATTCTTCTT	TCCATAATTG
156281	ATAGATCAGG	CAGACCAAAA	GAAAGAAATA	AGGGAAGATA	CGGAAGGTCT	GAACAATCTA
156341	AGAAGCGCAA	TCTCATAGTC	AATACATAAA	GCTCAGCAAT	TGTTTAATAA	TAGTAAGCAG
156401	AGAATATGCA	GTTTTCTCAG	GTATAGATGG	AACATGCACT	AACTGAGTAA	ATACTAGGCA
156461	GAAAACAGTC	TGAACAAGTT	TCAATAAATC	TGTATTACAC	AGATCATTTT	CTCTAGCCTC
156521	AATATAAGAT	TATAAACCAA	TAATAAAAAG	ATGACTAAAA	AGATTCTAAA	TATTAGGAAA
156581	TGTAAACTAC	TAATAAGTCA	TTAGAAGATG	TATAGAATGG	AACAATAATA	AAATGTTATT
156641	TATAAAATA	TACAATGAAG	CTAAGCAGA	ATTTTAAGGA	AAATTTGTAG	GCTTTAAATG
156701	CTTATCTTAG	AAAAATTAAA	AAGCTGAACA	TTAATGAGCC	AAGCATCTAA	TTTAAATTTT
156761	AAAAAGAACA	TAGAAAGCCA	AATATAATTT	TTTAAAAAGA	AAAAATAGAT	ATTAAACAAT
156821	ATAACAGTGA	AGTTAAAGAA	AACAAGAATG	CAATAAAGAG	GAAAAACAAA	CAAAAAAATA
156881	AGTAGCTTCT	TTTAAAGAA	ATTTAATAAA	ATAGACATAC	CTCCAATGAG	ATTTATCAAA
156941	GTAAGACAGA	AGGCACAAAT	GGAATGAATA	CAGAAACTTT	TAAATATTA	CAGAACTTTA
157001	TAATAAATCT	TATGCTACTA	ATAAAATTGA	AAGTACTGAT	AAAATTATTA	CTTCCTAGAA
157061	AAAATATTTT	TGAGTAAAC	TCACTCAAAA	AACAAATAAA	GCATGGGCAG	ACCTAACATT
157121	AAAGAAATGA	AATCACTACT	TTAAATTTTA	CCGACAGATA	ATAAAACGTG	CATCTTTATC
157181	AAGCAAAAAT	GGAACCTGTC	AGTTTTATAG	GAAATTTAGA	AGTCAAGGCA	TGAGTAATGC
157241	CAATCTCATA	CCAAATCCTA	CAAAGAATAG	AAAATTATGG	CTCCCGCTTA	TAGACATAGA
157301	TATAGAACTC	CTGCACAAAA	TAATATAAAT	AACAAACCAA	ATTTTATATT	TGCAACTATA
157361	CATATTATAT	GTGTATGTAT	TATATATGTT	AACATATACA	TATATAATAT	GTATAGCATA
157421	TGTTCTACAT	ATTATATATG	TATAGTGTAT	GTATTTTACA	ATATATAAAT	GAAAACCCAA
157481	TCTTTAATAT	ATTCATCTAG	ATTGTCATAT	ATGACATATA	TAATACATTA	CATCAAAAAT
157541	GTGTACAATA	ATCAGGCCAG	GCACAGTGAC	TCATGCCTGT	AATCCCAGCA	CGTTGGGAGG
157601	CTGAGGCGGG	TCAATCACTT	GAGTCCAAGA	GTTTGAGACC	AGCCTGGTCA	ATATGGCCAA
157661	ATTCCATCTC	TACAAAAAAT	ATGAAAAAT	ATCCAGGCAT	TGTGGTGCAC	ACCAATAGTC
157721	CCAGCTACTC	GGGAAGCTGA	GGTGAGAGGA	TCACCTAAGC	CTGGGAGGTG	GAGATTGCAG
157781	TGAGTCGAGA	TTGCGCCAGT	GCCTCCAGC	CTGGGTGGCA	AAGGGAGACC	CTGTCTCAAA
157841	AAAAAATTAA	AAAATTAGCC	AGGTATGGTG	GCCTGTTCTT	GTAGTCCCAG	CAACTGGGGA
157901	GGCTGAGGTG	AGAAGATCAC	TTTAGCTCAG	GTGGTGGAGC	CATGATCGCA	CCACTGTACC
157961	ACTCGGCTTG	GGCAACAGAG	TGAGAGCCTG	TCTCGAAAAA	ACAAATATAT	ACACACAGTA
158021	ATCAATATAT	ATATTATATG	TACCAATCAA	TGCTTCACTT	TTATATATAA	TATAGATTAC
158081	ATCTTATTAG	ATATATAGTA	TTCTTCTTCC	ATAGATAGAT	AGATACAGAT	ATAGACATAG
158141	TATCCTCTAT	CCATATTAGA	GAGAGGATAC	TATATATATC	TATAGCATAT	AGAGATGCTG
158201	TCTCAAAAAA	ATTTAAACAT	CAGCCAGATG	TGGTGGCCCA	TGCCTGTAGT	CCCAGCTACT
158261	GGGGAGGCTG	AAATGAGAGG	ATTGCCATTG	ATCCTCTCAT	TGGTTGAGCC	ATAATCGCAC
158321	TACTGCACCA	CTCAGCCTGG	GAGACAGAGG	GAGACCTGAG	GTGGAAGGAT	ATAGATATAG
158381	ATATATAAAT	AAATATGTAT	AGAGAGAATA	TAATATATGT	GTGTATGTGT	ATATATATAT
158441	ATTATGAAGA	CACTGGGAGA	GAATACTATA	TATATATGTG	TGTGTGTATA	TATATATTAT
158501	GAAGACACTG	GTGGGATGGT	TTCATTACCA	ATTGGACCAA	GAGTCCAGGT	ATGGAGCCAA
158561	CATGCAATGT	TGTTGTTGAC	TGAGCTGGCA	GAGCACTGGT	CATAGTTACG	GGAAAAGAAG
158621	GTCTCCAATG	AGACATACTT	AACAAAATAT	ATGAACTTGC	CATATACGTG	GAGAGTTCTG
158681	GTGTGTATAT	AGCCTTCTCT	CACCAACCTA	GCAATTGTCT	TCATCATCAT	TATAATGCTA
158741	TCAGAGCAAA	GATGACAGCT	AAATTTTTTT	GTCCCTTTCT	TCTTCTTTCT	CTTCCTTCCC
158801	CTCCCCCACC	TCTTTCTCTT	CCTCCTCCTC	CTTCATCTCT	CTTCTTTTTT	TTTTTGAGAT

Figure 9 (Page 49 of 74)

SUBSTITUTE SHEET (RULE 26)

138/162

158861 GGAGTCTTAC TCTGTCGCTC AAGCTGGAGT GCAGTGGCAC AATCTCAGCT CACTGCAACC
158921 TCTGCCTTCT GGGTTCAAGC AATTCTGCCT AAGCCTCCAG AGTAGCTAGG ACTGCAAGTG
158981 CACACCACCA CACCTGGCTA ATTTTGTAT TTTTAGTAGA GATAGGGTTT CACAATGCTG
159041 GCCAGGCTGG TCTCAAACCTC CTGCCCTCAA GTGATCCTCC TGCCTCGGCC TCCCAATGTG
159101 CTGGGATTAC AGGCGTAAGC CACTGTACCC GGCCTCCTCC TTTAATAGAC AGGGTCTAGC
159161 TCTGTTGCCC AGGCTGGGTA CAGTGGCGTG ATCATAGCTT ACTGCAGCCT CGAACTCCTG
159221 GGCTCAGGAG ATCCTCCTGC CCTAGTCTCC CCAGTAGCTG GAACTACAGG CATAGCACAC
159281 GGGGCTAATA AAATTAATTA GGTGATAAAA TTCAGTGCCC ACTGATGACT AAGCTCTTTG
159341 GACATAAAG ACACAGACCT TGAAGGAAAA TGTGTCTACT TAATTTTGAA ACCCTATTTA
159401 TCAAAAAACA GGATGAAAT GCAAAATGCC ATCCACATGC CAGAAGATAT CAGCTATAAT
159461 AAGTTCCCAT AAATCAATAA GGAAAAGAAC CCAATAAAAA TTATTAAGCC ACAGTAAATC
159521 ATGGGTAAAT CACAGAGGCC TGAAGGGCTA ATGGACATAC AAAAAGAATC TCAATCTCAC
159581 TAGTGAATC AGAAAAGCAC AAATTAAGTA CACAATTAGG TACCATTTTA AATCTGTAAG
159641 ACTGTCAAAA TCATAAATTA TATAAGTAA GACTCAGGGA GTTTTGGAGG AGTGAGAGCT
159701 CTTATATTGC TTGTGGGGTA GAATTGGAAC AATTTCAAGA TCTGTAGTAT CTGGTAAAT
159761 TATGATATGC ATCCCTCACA CCAGCATGTC ACTCCAAGGT ATCTCCCTGG AGGGAACATT
159821 TACGGGACAC AAGGAAGCAT GGATAAGAA GTTCACAGTA GTATTGTCTG CAACAGCAAC
159881 AACAACAAAA AAACCCAACT ACACACAAC TCAATGCCCA GTCCACAAGG CAATGGATTA
159941 AATAAACTTC AGGCCGGAGA TGGTGGTTCA TGCCTGTAAT CCCAACACTT TAGAAGGCCG
160001 AGGCGAGAGG ACTGCTTGAG CCCAGGAGTT CAAGACCAGC CTGAACAAAA TAAAGAGATA
160061 GTGTTTCTAC AAAAAATTTT TAAAAATTA GCCAGACGTG GCAGTGCTTG CCTGTGGTCC
160121 CAGCTACTGG GGAAGCTGAC GTGGGAGGAT TGCTTAAGCC CAGGAATTTA AGGCTGCAGG
160181 GAGCCATGAT GGGGCCATTG CACTCCAGCC TGGGTGACAG AGTGAGACCC TGTCTAAAG
160241 AGATAAGTAA ATAACAATT TGCATTTTCT GCCACATTGC AAAATGGTGA GAGAGTGGTT
160301 TCTAGACTCT AGACTCTTCT TATGACTACC TTCTAGTTAT GAGATCCTAC AACACTCACC
160361 TAACCTCTCT GTGTCATATT TCCTCCTCTA TAAAGCAAAA ATGCCCCATA TAGAGAGGAC
160421 TGTGATATAA AACAAGAACC AAGAAAAGTA AAGCTTTTCT AATCTGTCAC AGACTAAAGA
160481 GTGCTCAGTA TATGTGAGTC ATTATTCCTG GTGCTGGTAG GAGTGTATGT TACAACCTTG
160541 AGTCAAGTAA TATGGTACCA TATATTAAGA TTAACAACAA CCTCGGCAAT CCCAGTTTGG
160601 GGTATGTTCC CAAAAGAAAT GAAAGCACCA GGATATAAGG ATGCATGGAC TAGAAAGTTA
160661 TTGTAGCAAC ATTGTAATAA CTAAGTTCTA AAAACAGCCT GAAGCTCCAT CAGTAGGGAT
160721 ATGGTTACAT ATATTTATTA TATTCTTATG GAATATTAGA CATAAAAAAGT AACGAGTAAC
160781 ATAGAAGAGA CAGTGTATAT ATGTTACGTT TGTACAACT TAGGGAAGA TATAGATCAC
160841 CCTACCTAGA GAAGTCAGAT TGGAGAGGGG TGGGAAAAAC CTTGAACCTT CTCCTTATAT
160901 CCTTTATATT GTTTGACTGA TTAAATGTAT TTTGTTGCAT CTGCTTGAAG GCAATGTAAA
160961 ATAAATAAAA CATACTTTA. AAAATAAAAA TAAAATTTAT TCCTATCACT TTTGTATAA
161021 AGCTGGGCAC AGTGACTAAC ACTTGTAATC CTAGCACTTT GGGAGGCAGA GACAGGCAGA
161081 TCACCTGAGG TCAGGGGTTT GAGACCAGCC TGGCCAACAT TGTGAAACCC CATCTCTACT
161141 AAAAAATACAA AAATCAGCCA GGCATAGTGG TGCCTACCTG TAATCCCACG CTACCCGGGA
161201 GGCTGAGGCG CTGGAACCCA GGAGGCAGAG GCTGCAGTGA GCTGAGATTG CGGCCTGCA
161261 AGCCAGCCTG GGTAACAGCG AGACTCCATC TCAAAAAAAA ATTTGAAAAA AGAAAAATTT
161321 TAATAACAG TGTTAAGAG GGGAGAAATA TTTAGTTAAA AGATAAGCCC ATTTAAGAAA
161381 TAGTTTCACT TGACCCGGA GGCAGGACTT GCAGTGAGCC GAGATCGCAC CACTGCACTC
161441 CAGCCTGGGC GACAGAGCGA GACTCTGTCT CAAAAAAGAA AAAAAAGAAA GAAAGAAAGA
161501 AAGAAATAGT TTCACCTGAA CCATATTATG ATTCCTTCTG TAAAAGATGA GAGTAGGCCAA
161561 ATTGACTCAG TGAAATCCCA GCAAACTTA CACAAAGTCT TGTTCTTCTC TCCTGTCATC
161621 TGTATAGGAT GAAATACAGA GTGCTTTTGG GTTTTGTGTT TGTGTTGTTG TGTGTTTGTG
161681 AGGGGAACAC AGGTCTATAA TTCCTTTTCT GAAATCCCTG GAACAAAATG GGCTTTGCCA
161741 TTCAAATTAG TTTAGAAGTT ATAAAGGCAA AAAAATGCAT ATACTCTAAA GTTCAACCCC
161801 ATCATGGCCT AAGGCAGAGC CCTGTAATCA AATTCATCAA TATATCTGCA GCAAAACATT
161861 TATTCAAATT AAGTGGGATA AATAAGACT TTTAAATAGT CTCATCTCAG TGCCGTTTCTG
161921 GGTGAGCCAC TGTGGAAGAC AGACTCAAGG GTGGCCTTCT ATGATTCCTG CCTCTTGGTG
161981 TTCACACCCT CGTAAATTC CTTGTCTTTG AGTGTGAGCA GGGCTTATGA ATTGCTTCTG
162041 ACCAATAGGA TATGGCAAAG ATGATGGGAT ATAATTTCTA TGATTACGTT TCATTATGTA

Figure 9 (Page 50 of 74)

139/162

162101	AGACTCCATC	TTGCTGGCAG	ATTTTCTCTA	AAGAGTCTGT	CTCCTGAGCT	CTCTCTGAAG
162161	AAATAACTGG	CCATGTTAGA	AGCCCATGTG	CAAAGAGCTG	AGGGGTGGCC	TGTAGAAGCT
162221	GTGGGCAACC	TCCAGCCAAC	AGCCAGAAAT	AACCAGGGCC	AAAGTCTGTC	AACCATCAGG
162281	AAAGAAATTC	TGCCTGCTAT	CTCAGTGAGC	TTGGAAAGTGG	ATTCTTCCTT	AGCCTAGCCT
162341	CCAGATAAGA	ACACAGCCTG	ACCAACACCT	TAACTGCAGC	CTTATCAGAC	CCTAAGCAGC
162401	AGGCCCAACT	AAGCTGTGCC	CAGATTCCCTG	AACCACAAAA	ATTGAGATAA	CATATCAGTG
162361	TTGTATTAAG	GTTCTAAATT	ATGGTAATTT	GTTTGTACTA	ATAGATAACT	AATATAACCA
162421	CCAAATCATT	TCAGGTTAGG	CCAGATTTTT	GTAGCCAAAT	GAATCATGAT	AAAACTTTCC
162481	ATTTTCAGGG	GTTTTTTTGA	TTTTGTACTT	ACGGATACAA	ATTTGTGAAA	GTATAGTCAG
162541	CAGTGATTTA	AAAAATCAAG	GGAGCAGGAA	ACTCAGTAAA	TGGTTCTAAC	ATTTTGGAAAT
162601	CTGTAAATTG	GTTGTAACAT	TTGTCACTGT	TGTTATCTAA	GTCAAGTTCC	TAAAAATATGT
162661	GAATGATAGG	TTATCATACT	CACCTACTTT	TCTTGCATTG	CTCTAAGAGT	TGGCTGAGCT
162721	ATTGATAATA	AACACTATGA	TCAGATCTAA	TACCATGATG	TGCTATTATG	ATCATGTGTC
162781	AGTCACAGGG	CTAAGCACTT	TGTACATGTT	GATGCATTTA	ATTTTGATGA	TAACTCAATG
162841	AAGTAGGAGC	TGTTAATATT	TTCATTTTTT	AGAGGGGGAA	ACCAAGTCAC	TTGGAGTAAAC
162901	ATGGCTAATA	AGTGAAAGAA	TAAGAATTTG	AAAGGTTTGC	ACAGATAACC	AGAATGCAAT
162961	GCTCATCACA	TTCAGTGAGC	AGTGAATCAT	ACTAACTAGA	GAAAGTATGA	AAGCTCTACT
163021	GAAATTAACT	AAACAACCTC	TCTGGCTGTG	AGCCTGCCAA	GGGACAGGTG	GTAAACTTGG
163081	TTACTGCATA	AGGCCCTTTC	TATCCACAGT	ATTCAGGAAT	TCTTTAGTGA	ACATACCTTG
163141	ATGACTCCTT	AACATTTTCT	TCACATCGAA	GTAAAGCTTG	GAAACATTGC	ACATAGTATG
163201	AAGTTCCAAG	GAGACAGCCT	CTGATGTTTT	CAGCTTCACA	GCCCAACTCC	TAGAATAAGC
163261	AGAGGCGAGA	GATTTCTTCA	GAGGTGCATT	CCATTCATTT	CTATATACGC	ACACCCCTCC
163321	CCTCCTGCAT	TCAAACAGGA	CTTACCTGCT	CAAAGTGTC	TTCACATTCT	ATAAAGAAAC
163381	AAAAAGAAAA	GGTGAGCATG	GGAACATCGG	TATTTTCATG	GGCTTGTCAT	GCAGGGCTAT
163441	TCTTCTTTGC	TTTACCCGAA	GAAGTAAAGA	GAGTTACCCT	AGTCTTAGTC	TTAGATATTG
163501	ATGGATACTC	AAACAAAGTA	ATTTCCACCA	GTCTTAGGTA	TTGATGGATA	CCCAGATGGA
163561	ATAATTCCTA	CCAGCTTCTG	GGAGATTTCAG	CATGGCAGGA	TGTTTATCAA	CATTTGCATC
163621	TATTCTCATC	CTTGCTGAAG	TCTGAGGGCC	AGGAGCTTTG	TCCATGCTCC	CTCTGTAAAG
163681	ACTAGCTTTT	GGTGATCGGA	TTTCTTTCAC	AGTGAGCCCA	GATTAGAGAA	CACCTTATCAT
163741	AAAGGTCCTT	AGTGGTGAAT	CTGTGCACAG	CCCTGAGACT	GGGCCACTGC	CACCTAAGATG
163801	GTGGTAGCAG	GTATCACACA	GTGGTAAAGC	AATCATGCTA	TACACTCAGC	CTTACAGTAT
163861	AGTCACCAAT	CCTGTTAGTT	AGAACCAGAA	TTAATGGCTC	CAGATGTTTA	TCTTCTTACA
163921	GATAAAGCTG	TAGATTGTAC	CATAACAGCT	CTGGAGCAAG	GTTTCTACAA	GCAAAATCAGG
163981	GAAAAGGTTA	TCACTCATTT	TGGCTGCCCC	ACTTCATCAC	CCATCAGTCA	CCTAGTGGAG
164041	TATTTTCAGGA	GAGAGTCAAC	AACCAGGGTT	CTCTGCACAT	GGGCCAAGGA	GGCAAACAGT
164101	GGTAAATGTT	ATCCCGTGGT	TTCATTGGCC	CAAGCTGTGT	TCCCTCAGAA	GTTTATTTTT
164161	CTAATTGACA	TAAAGGTACC	CTATAAATTA	GTGAAGGCCA	GCCTGATGGC	ACTGATGTAC
164221	ATCTAAAAGA	AACATTACTT	TATCTTCCCA	TGCTTCTTTA	CCATTCTCCT	TTAATAGCAC
164281	TATAACATAC	CTTTTTTCCC	TACTCCAAGT	ACACAGCCTC	ACCTGCAGCA	ATTTCTGGGC
164341	TGAGCCCTGA	CATTTTTTCT	CCAGTTCCAG	GATGTGGCTC	TTGAGTTCAT	TGCTCTTCAG
164401	CCCCAGACCA	GCCTCATAGT	CCCTCAGTCT	ACTCAGAGTC	TGTTGTTCTT	CTTTCTCCAG
164461	CCTCCAGAGA	TAAGACTTCT	CTTCTCATG	TAGGAAACAC	TGGAGATTCT	TAAAGTCAGA
164521	CCGGATTTTT	TGTCTCTGAA	TCTGTACCTT	CTCCTGGAGT	CAAGAAAGTA	TGGTCAAAAG
164581	GTGGAAGTAA	ACCAAATGTC	CATCTATGGA	TGAATGGATA	AACAAGAATG	AAAGTCTGAC
164641	ACACGCTACT	ACATGACAAG	CCTTGAAGAC	ATTCAAGCAA	AATAAGCCAG	AAACAAAAGG
164701	GCAAATATTG	TAAGACTTTG	CTTATACAAG	GCATCTGGAG	TAGTTAAGTT	CATAGAGACA
164761	GAAAGTAAAA	TAGTGGTTAC	AAGGTGTTGG	CAAGACCAGA	AAATGGACAG	TTATTGTTTA
164821	ATGGGTAGTG	AGTTTCAGTT	TAGAAGATGA	AAGATGAAAC	TGAGTTGCAG	TTTGGAGATG
164881	GGAATGGTGA	TGGTTGCACA	ACAATGTAAC	AATGTAAAAG	CACCTAATTC	TACTGAACTA
164941	TATACTTAAA	AGTGGTTAAA	TGCTTAAGTG	TTATATATAT	TTTCACACAA	ACACACACAC
165001	ACACACAATC	AGCCACTGGG	ACATTATTTT	CTCATGAGTC	ACTGAAGCTG	GAAGAATGTC
165061	CCCAGTTTCC	TGCTGCAGAG	TCATGTGTGG	GAGGCAGGCA	CTCAGATGTG	GAAGAGGTTG
165121	CCTCAGATTG	CTTATAGTCA	CCCAATTAAT	TTTCTTGTTT	TTTCAAGCAAG	ACACAGGAGA
165181	AAGCTGGGTT	AGGAGTGCTA	GATAATTTAA	TTGTGAAACT	AGGGCCAAGT	TCAAACACTT

Figure 9 (Page 51 of 74)

SUBSTITUTE SHEET (RULE 26)

140/162

```

165241  TATCAGTTAC AAGGATAAAA AGAGGTTTTT ACTTATGATT TAAGAAGTTA GATTTCTGAG
165301  TTGGAGCGAT TTTCTTGAAG TAAAAGCTTA TAATGAACAT CACCCAGACT GGATTTTAAG
165361  ACAACCAGGC TGTAAGAGG GTCCATAATT CTGGCAGGG GGAGCTTTGA GTGTGACAGG
165421  CATTTATTAT GGTTAACTGA GAAATACTGT TCTACTACCC TAGGGTCATC TTAAGCATTC
165481  CTATGTGTAA GACTGACAGA AATCAAGTGA AACTCTCATC TGAGGAGATG TAAAGTTGCA
165541  ATTTCCATTA GTGCTGTCTA AATTAATGCA GTGGGAGTGT GTATTGAGG CAATTTGAAT
165601  CTATGTTCTT GGATTGCAGT CTTCAAACCT GGCCCAAATA AACTCTCTAC TTATCTTAAA
165661  AAAATAAAAA TTAATAAATA AAAATAAATT CATACAGTGT TTTGATGACT ATGATATAGA
165721  AGAAGGGTCT TTGACTTAGG ATGAGGTGGA ATTTTGTGT AGGAGACAGG TGCAGCTTTA
165781  ACTCTGTAT AGACGGGTTT TCATATATGT TAGTTACAAT CAAGGTCTTC CCCATTGCCC
165841  AAGATCCTAG AAATGGGGGA AGTAAGAGT TACTCAGGAG CTCAAGAGCA ACATCCACAA
165901  ACAAAGATCA GGGTAGAGGT TAGAGAGGAC TCCTGAAAGA GAGAAAATTG GTAATCAGCT
165961  TGTGGGATTT TACTGCAAGC TAGTGAATTA TATAAATATA AAGATTGGTG CAAAAGTAAT
166021  TGTGGTTTTT GCCTTTACTT TAATGGCAAA GACCGCAATT ACTTTTGAC AAACCTAAAT
166081  ATTTCCATAA AAGAATGTGG CTCTGATAAT GTGGAGGTTA GTCAGCCACG GAAATAATCT
166141  GAAAGTTTGT AGTTGCAAGT GTGTAGGTTG TTGCATTACT TGTGATGTAC TTATAAATCA
166201  AGTATAGGCC GGGTGACAGT GCTCACGCCT GTAATCCCAG CACTTTGGGA GGCTGAGGTG
166261  GGTGAATCAC GAGGTCAGGA GATCAAGACC ATCCTGGCCA ACATGGTGAA ACCCGTCTC
166321  TACTAAAATA CAAAAAATTA GCCAGGCATG GTAGCACATG CCTGTAATCC CAGCTACTCA
166381  AGAGCTGAG GCAGGGGAAT GCTTTGAACC CGGAGGTGG ACATTGCAGT GAGCTGAGAT
166441  CGCACCCTA CACTCCAGCA AGACTCCATT TCAAAAATA GTAATAATTT AAAATAAAT
166501  AAATAAATA AGTATATTTT TTTATCAGC TTCATGAGCT TGAGTAGTAT GAATTTCAAT
166561  CTGGAGTGAT CCTGTTTTCT AAGTGTTTAC AAAGCTTGGT TTCTGTACCT GTAAAGTTGA
166621  GAGCCAGATG CTCCACTGTG GTAAAAGTGC CAGGGTAATG AGTTGAGGCC TGCAAACCAG
166681  GTTTATTTTG AGGTATTTAA AGTTTGAGAC CCACTCGATG CTTTTCTAG GTAAATAGTC
166741  ATACTAATTC TGCTTCTTCT GACTGAAGTA TCAGGAATCC CAGCCAATA CAGTTTAAAG
166801  ATGGAAGAT TGGTGCTAAA TACTCATGGA TGTAAACCTG GAACCAGGGG CATAAGTACA
166861  AATAATGGTT TCTTCCTTGG GTTTCATTTT TTCAATCTGG TTTAGTGAGA ATAAATCCTC
166921  ATGTGTCTT TCCTCAATCA TCCCCTATGC CTAAGCTCTA GAATGGAAA TAGCTTGAGA
166981  TCAATGAAGT CAGATTCTTA CTTTCCATT AGTTATTCGC ATTGCTGTGG ACAGCTTCTG
167041  CTCCGTACAT CTGTCTTCAA GTTGCTTCAG TTTTGTACA GCTTTCTGGA GCTTTCTCTG
167101  AAGGAAAAAT TTGATAAGTG AAGCCTATTC AATTTGACTC TTCATTAGGG ACCTAGGGGG
167161  AATCCCAATC TTCTAAGATA TATTTGAATA ATAGTGAATA TTTATAGAGT CCTCATTGTT
167221  TTTTGCTAGA GAGCATGCTA AAGGCTATAT GTGCAGGAAC ATACTGATCC CCTTGGCAAC
167281  CCTGAATAGT TGGTAGGATT TTAAACTTCA TTTCTGTGCT GTAGAAAATG AGACTAAGAA
167341  AGGGGTAATA TAACTTGCCC AAAGGGCTAT GACTGCCAGG TGGTGGAGCA ACAATTGCAA
167401  TCTCATCTGC TGACCCAGAG CCTGAGCTAT GTCCACCACT AGAGTCCTGC CAGGAAAAAG
167461  TTGGATATAG AACAAGGTAA TCATCATCTA AAAGATTTTG TAAACAACA TGCTGAACCA
167521  AGCAAAACCA ATACCAGTGT TTGGCACACA TGAAATTTG TGTCTTATGA GTCAGGAAAA
167581  ATCAGGATGC CAGCTGGTTA TTAGAACAG TTCATGGAAG AGGGGAATTC TGGTATCTTT
167641  TGAACAATGG TATCATGAAT CCAATTTAAA ATGATTTAGT ATTCATGTCA AGCTTTTAGC
167701  TTATTTCTCA AAACAGTTTC TCATATTTCT ATTGAAAGTG ATTTGAAGCT GACCCAAAT
167761  GCTAATTGTA GTCAATGCTG AAAGAATTGT CTCCTGTCCT CTGTAAACCC AACAAGTATA
167821  CTCATTCAAT CTCGAGTGTT CTCAGGAAAA GGTTCATATG AACTGTTTTA GCAAAAGATG
167881  ACATTGTCCT TACTATATGC CAAGTGCTAT TCTATGCATT CTATATTTTA ATGTCCTCAA
167941  AGCTTATAAC CACCTCCTGT GTATGTGTTT TAGGGAGGGA GGACACTGCT ATTATCCCCA
168001  TTTACAGATG GAGAAACCAA GGTGTGAAGA CATTAAGTAA CGTGCCCAAA ATTGCCATC
168061  TAGTAAGTGA CAAAACCTCA TTTCAACATA AGCTGTTTCC TTTTCTTACT ACTGGGTGGA
168121  AAAGTAATTC AAATGGGAAT ATGATCATCG CAGTTATTAG CTGCTCCATG GAGTTTAAAG
168181  AAGAGCTGCC ATGAGCTGAG TGGTGGTCAT GATTGACATG TCCTTAGAAG GACTTAGAGC
168241  CTTATACAA GACCACCTCT GCCTCATGGA GGACAGAATA AGGAGCCTGA CACTGGAGAC
168301  AACATTTTCC TCAAATTTAG GCAGGACAGA GAAGGAAAAA GGACATCAGG ACTATGCCCA
168361  TTCTCCATG CTGCCAACAG CAAAGTCCCA CCTCCTTAA TATGCTTTCT GGCAAGAAAT
168421  CTGGATGGTA CACAAACCT CTCCTCTGCT TACACCTTC ACAACCAAGC ATTTCCAAAT

```

Figure 9 (Page 52 of 74)

SUBSTITUTE SHEET (RULE 26)

141/162

```

168481 CTTTGA CTCTCT TCTTCCTGAA TCGTGCTTAA AATCTGCCCT CTCCTCCCTT TCTTATACGG
168541 ATAGTTTGAA TTTTACTCCT TGATATTCCT TTTATCATAG ACATGCCACA GTAGCTGGGC
168601 ACAGTGGTTC ATGCCTCTAA TCCCAGCATT TTGGGAGGCT GAGATGGGAG GGAGACCAGG
168661 GGTTTGAGGC CAGTATAAGC AAGAAAGGCA GACCATGTCT CTACAAAAAA TAAAAAATT
168721 ATCCAGGTAT GGTGGGGCAT CCCTGTAGTC CTAGCTACTT GGGAGGCTGA GGTGGGAGGA
168781 TTGCTTGAGC CCCAGAAGGT TGAGGCTGCA GTGAGCCGAG ATTGCACCAT TGTACTCCAA
168841 CCTGGGATAC AGAGCAAGAC CCTACCTCAG AAAAAAAAAA AAAAAAAAAA AAAGTAGAGG
168901 TACCAGAGTG ATATTTTCAA TGTCAGTGAC CCTTCATTCC CCAAATGAAA ATCCCCCAAT
168961 AGGTGTTCAA TTTTACGTG TCCTTCAGGA GTTACTTCTA AGATGAACCA CTCTCTACCC
169021 TAAATGTCCC TCCCCACCAC CAAAACCAGG GACCTCCAGG CAGACATTTT TGATGGTTTG
169081 TTTTCTTTAC TAGACTGTAG ATACCTAAAA GGTGATGGGT CTTTCTTCCC TGTTTTCAGG
169141 CCCTACTGCA TGGCTTTTACA TATTGTGGTT TTTCAAATGA TATTCATGGT GTGAAACAAG
169201 AAAAAATGCG GGTGTTTGGT TTGAGAACA CCGTGTCTAA AGCAAAAAAG AATTCATCAT
169261 AACACAAATG GATAGAGATA AGAGTCCAAC CATCCCATTTG AAGGTCAGGA TGGACGCTCT
169321 AGATAATTGA GCAAGAAATC ATCATAAACT ATTTTTCAGA AGAATGACAT GATGAAAGCT
169381 GTATTTCCAA GTCATAATGT TAGGTTTCAA GTTAAATCAT CTCAGCTCCT GGGGAGCAGG
169441 ATAAGACTTG GTACTTACCA AAGCTCCCGG GCCCACACAC TCACCTTGTA GCCCTGGCAT
169501 ACGTCTTCAA CAAGAGCTGT GGTGTGCCCT TTGTGCTGTG GTGCCCGCTC ACAGCGCCAG
169561 CAGATGAGCT GCCCCTCATC TTCGCAGAAC AGGTGGAAC TCTCTCCGTG TTCTCTACAT
169621 GACATTCTT GATCCGTCTC TTTGAGGGCT TCAATGAGGC TTCCAGCTG CTGTGTGGGT
169681 CGGAGGCTAT CCATATGAAA TGGAGCCGGA CACTGGGGAC AGCAGAATGT CTCCTGCCTC
169741 AGTTGCTTTT GGCTTGGGTT TTTAAAGAAG TCTGTTATAC ACAAGTGGCA GTAGCTGTGT
169801 CCACAGTTGA TGCTTACTGG GTTCGTCATC AGGCTCAGGC AGATGGAGCA GGTGGCTTCC
169861 TCCATCATCT TCTTGGTGCT GGTGGTTGAG GCCATAGCTT TTATTGAAAA GCTCCAATAT
169921 TGGCTCTAGA GATGGAGATG AAGCAGCCAG AATTTTCCAC CGTGATGAAA ATACACCTCA
169981 CCTGCACCTC TATGTGATGA GCTGGCTGCA ACTGACTTCC ATAGGTCTTG AAGGTTTTCC
170041 TTCCAACCCC TATTATCTCA TTTTGTATG AAGAAAAGAG GACCTAAAAG GAAGAAGTTG
170101 AGGCTGAGGT TGTTTGGGCC ACGTTTGAGA ACTGCAACCC AAGTGCAGAG TTTCAAGTTG
170161 CCCTCATTAG CAAGCAGTTA CAAGTGGTTG TTTAGAGGAA AAAAAGCAGT TTTAAAGCAG
170221 TTTTAAAGTT GTTTGCCAAG AATTACATT AAAATAGCAT AAGCTTTTGA CTGGCTATAC
170281 ATTGTTCTTT GTATTACAAA TCTCGGGAAT ATGTAGGTAA TAGATGAGGC AGCCACTCAG
170341 GAACAAAATG CTTTTAAACA TGGGGTCTTA ACTGAAGACC TATACTCCTG CCTCACTTGT
170401 CCTGATAAAT TTTGCATACC TCACATAGCT CAGACTGCTC TAAATTATTT CATTATTTTT
170461 CTTTTCTCAG TCTTCTAACT TTTTTTTTTT TTTTAAATGA GACGGAGTCT CACTCTGTCA
170521 CCCAGGCTGG AGTGCAAGTA CGCTATCTCG GCTCACTGCA CCTCCGCCTC CCGGGTTCAA
170581 GCGATTCTCC TGCCTCAGCC TCCCGAGTAG TAGCTGGGTC TACAGGTGTG CACCACTACG
170641 CCCAGCTAAT TTTTGTATTT TTAGTAGAGA TGGGGTTTCA CCATGTTGGT TGGCTAGGAT
170701 GGTCTCGATC TCTCGACCTT GTGATCCACC CGCCTCAGCC TCCCAAAGTG CCAGGATTAC
170761 AGGCATGAGC CACCGTGCCC AGCCTCTTTT TCTTTTCTTA TAAGACAAGT TCTCGCTCTC
170821 TTGCCCAGGC TGTAAGTGGG GGCAGTGGCA TGACCACAGC TCACTGCAGC CTCGACCTCC
170881 TGGGTTTAAG CAATCCTCCT GCCTCACCTT GGCAGAGTGG CTGGGACTAC AGGTATGTGC
170941 CACCATGTCC AGCTAAAGTC TTCTCTCCAG AAAGAAGAAA TGCATTGGAA TTTAGAGGAT
171001 ACACAAACAT CTAGCTGTAT AGCTAATACA GTAGCCACTA TCATGAGTAG GAATTTAAAT
171061 TTAACCTAAT AAAAATTAAA ATGAAAAAAT TCAGTTTTC TGTTCCAGTT GCCACATTTT
171121 GATTGCTTAA TAGTTGCATG TGACTAGTGG CTACATAACA GCCTCAATAT ACAACATTCT
171181 GTTATCACAG AAAGTTACCT TGGACCAAGT GCTGGGAGAA GCAATGCAGG CTTCCTCACA
171241 AAAGCTGTAA AAGAGAGAAC TCAGGGAGTG TGAACTCTT TCCTATTCTA GTTAACTTCA
171301 AGAATAATTG TTACCAGGCC AGCACGGTGG CTCACGCTG TAATCCTAGC ACTTTGGGAA
171361 GCCGAGGCGG GCAGATCACC TGAGGTCAGG AGTTTGAGAC CAGCCTGACC AACATGGCAA
171421 AACCTCATCT CTAATAAAAA TACAAAAAGT TAGCTAGATG TGGTGGTGCA CACCTGTAAT
171481 CCAGCTGCT CAGGAGGCTG AGGAAGGAGA ATGACTTGAG CTCCGAGGG GGAGGTTGCA
171541 GTGAGCCCAG ATTACACCAC TGCATCCAG CCTGGGTGAA AGAGCGAGAA TCTGTCTTAA
171601 AAAAAAAAAA AAAAGAATAA TTGGTACCAG AATTACTCTT TGTAATTAGT AGTAACACTT
171661 ATGCAATTGG GTGATCTGTG ACAGATTCCA TTGAAGGAGT ATGGGGAGCT TCACCCCAAT

```

Figure 9 (Page 53 of 74)

SUBSTITUTE SHEET (RULE 26)

142/162

171721 ATATGACTCC CTGGTATAAT GAGTATTTTG AATTAAAGGC CCTTAGAGAT CAGCAGATGC
171781 TGAAGAGAC TTTCCCCTA TCTACATAAA GACCAGTCAC ACTAGACAAG AAGAACAATT
171841 GTTTTTCCTT CCAACCCCTA TTATCTCATT TTGTACTGAA GAAAAGAGGA CTAAGAATGT
171901 AACCAGACCT AATCAGACAC TTTCACAAA TAATGTCTGT CTCTCAGGCT CATTCAATTT
171961 CCAAAGAGAA CCATTTACAA GTTAAACTCT GTTCCTCCAT TCATTCATCC TCCCAAATAT
172021 TCATTTATTC TCCCTAGTAA TCATTTACTG CCCCTCAAAG AATTACCTAT ATTCTCCTGA
172081 TATCACCCTT CCCCTCTGAA ATAAATATGT ATACATGTAT AAACGTTATA CATACATATT
172141 TATACAGTAT ACATACATAT TTATACATAC ATACATATGC ATACATATTT ATATTTATGT
172201 ATTTATACAT AAGTATTTAT AAATAAGGCT ATATAAGTAT CTACCCCAT TGGCAGAGGG
172261 GGTAATCACT CTGTGATTCT AGCCCATGTA CTGTGTAATA AATTTGTATG CCTTTCTCC
172321 AATTAGCCTG CCTTTTGTGA GTCGATTTT CAGTGAACCT CAGAAGGCAA AGGGGAAGTG
172381 TTCCCTTGGC TCCTACACCA TCATGACAA AAAATTTGAC TCCACCTCGA CCCCCCAT
172441 CCCCCACAAA GAACAACAAC CAACACTGGT TAATAAGGTC GGTGTTTTT TGTGTTGTT
172501 TTTGTTGTTG TTGTTGTTGT TGTGTTTTT GCTTTCAGGA GCAGAGGTAT AATAGGCAAA
172561 AGAAAGAGAA AGGAGAATAG TGAATACCTC TTCTGCAGAG AGGGGTGCCT AAGTGGGACT
172621 TCCCTGGCTA ATAACGTCTT GCTAGAGACC CAACCAGGAG GATAATGGAA GCAATCAAGG
172681 CAACCAGAAC AACCAGAAGA ACCAGTTTAT CCTTTTGTG CCCTCTCCCT AAAGTGGGG
172741 AATAAGAATT GGAAAGAAGG CTGCAGAGCA GAGGGTTTGC TCCTGAGGAG CAGTTATTTT
172801 TATGGGATCA GAGCTCCTGC AGAAGTGGGG AGTTTACTTT TACTATCTCT TCTCCAGGAC
172861 AGGACCTATC TCAAGAGACA TGTTCCAGAG GATTGCAACA TAAAGAGTTT GCAGACCCAA
172921 GGAGGTAGGG AAGGCAGAAA GAAGATGGGG GAGGCCAGGG ATAGGCAACA GAGGAGTGAC
172981 CAGGAGCGAA AAAGCCTGCC TCTTCTGAGA ACCTAGCTGG GCTCTCCCTG TACCCCGAT
173041 CCCTCCCCC CGCCCGCCCC CACACCCCTA CTCCTGGGAG CTCCTCTAGG ACAGGGGAG
173101 AGTCAGGAGG AAGTTTGAAG AGTGCCTAGA ATAAAAACA GTAATTTAAC TACAATTACC
173161 GGGTAGGCTG TTTTCTCTC ACAATTTGAT CAGTCTCTG AAGCCACACA GAATTTCTTC
173221 TGAAGACGTG TATTCCTTGG CAGGCTATTT CCTCCAGTGA TACACCAGGC CCCTCTCTGC
173281 TGGGGTCACT GCTCTTCTGG GGAGATGGGG CTCCTCTCT TCCAAGGCTC CAGGGTTCTT
173341 GTCCTGGGCC CCACTCATCT AAGTTCTGAA TCTTCTGAGA TTTGGTGTAA AGTCTGGTGA
173401 AAGAAAGAGC AGGAAAGAGG TGAGAGCTGT AAAACAAAGA AAGTCTGAC CATTTTCAGA
173461 GTTGGAGGGG CCCTGCTGTC ACGAAATATA TTCCCACCC CACTTGCCAT CAGTACACAC
173521 TCACATATCC ACTGAGAAAA CCTTAGCCTG GACCTTTTCC GTAACCTTCA CTGCTCAGAC
173581 ACTTACATAT TCGCTGCTAG TCCCCTCTGT TGCTGCCACT TCCTGGGTCA GGAAGTTAAC
173641 TCAGACCGGA TTAACTGAG AAGTGAACT ACTGTGGGAG GCGGGGCTCA TAAGATTTAG
173701 GAGAAAACTA GTGACGTTGT TCATATCATT TGCACTCCGC CTCTCCGGTA AAGGAGGGGG
173761 AAACGTAGGA AGAAAATATC CTTCTTTTAC AGCAATAAAA AGAAGGAACC AATTAATAAC
173821 CCTGTAACT ATCATGTGAC CCCAACACAG AGTATCTAAA AACAGGAAGC CTGCAGAGGT
173881 TCAGTTTACA GACTCTGATT TGAGATCTTT CTACTTTTGC CACCAACTCC CTTGGGAGTC
173941 CTTAAGCCTT CTTAGCTGAT GTTACTTCTT TTGCTATTTA TGGGTTGCTT GTGGTTCTAT
174001 AACTGCTCTG AAGGGTGTGG TGGAAAAAGG GGTGGTAACA GCAGTAGGAC TCATTGGCAT
174061 CACAAAATTC ATCTGAGTCA GCTTCTATT CTTCTCTGTC CCGTTCTGTG TCTTGTTTTT
174121 CTCCTTGCTG TCCTTCTGCA GGACTCAGAT CTTCTTCAAT AGCGAGGGTC AGCCAGGATA
174181 GAAAATGGGA GTCAGTAGTG GCCCAGCAGT GAGTGCCCC AGCTTAGAGC TGTGTGGGAT
174241 CCCTGGGACC ATCACTCTGC TTTGTGCTTT GTGGAGAAAA GGCTGTGGGG TCCAGGGTCA
174301 AGTCCTTAAT GACTTAGCTC CAGCTTCTCC ACTTCAAAAT GAAAGGAAAA GTACTATCAC
174361 CACCCGTTAG AATTATTATT TCATGGGGAA AAAAGATGGA TTACTATCTC ACAATAAGAG
174421 CTTGTACAT TTATAAGTCT CAGGTGTAAG AGGCATTTAT GATAACAACA TAATAAATGC
174481 TGGCTTAAGT AGATGCAGTG GTCCAAGGGA ACCAGTAAGG GGAGCTCAGG ACACAGGTGG
174541 GAGGAGAAAT TAAACTTGAA TTCTGGGAGC CACTGGCCTG TCTGGGCCCC TGGCCTGCCT
174601 GCTGACCCTG ATAGCCAATG GAACATGGAG TTTGGCCCC CTGCAATCCC CTGGTCCAA
174661 CTACTCAAAA TAAAGGCAAG ATTGGGAAAC ACGTTCCTTT CTTCTATAC CAAGCAGAAG
174721 ACTCTTCAGC ACTGCACCCT CCTGGGTGCT CACAGAGCCT TCTGTTGTTT TGCCACCTAC
174781 GATTATCAT GGCCTGGCAT GATGGTTGCA GACCCCATGC ATAGCATGGG ACATTCTACT
174841 CCTGAGGCAA CCAGCACACA GAGAGAGGAG AAAGAATGAG CCCCTGAATC CTTGGTCCCA
174901 CGATGAGTCC TTGCAGATAT CTACAACCTT CATTGTTGTG GATGTGACTC TGTACCCAGG

Figure 9 (Page 54 of 74)

SUBSTITUTE SHEET (RULE 26)

143/162

```

174961 CATGGCTCAT TCCAGATCTG TCCTATTGTC AGAGGTGTTT AAACCAGAAAT GACTCCATTT
175021 TGAATGGGGG CTAGGTAAAA TAAGGCTGAG ACCTACTGGG CTGCATTCCC AGGAAGTTAG
175081 GCATTGTAAG TCACAGGATG AAATAGGCAG TTGGCACAAG ACACAGGTCA TAAAGATCTT
175141 GCTGATAAAA CAGGTTGCAG TAAAGAAGCT GACCAAAACC CACCAAAATC AAGATGGCAA
175201 CAAGAGTGGC CTCTAGTCAT TCTCATTGCT CATTATACAC GAATTATAAT GTGTTAGCAA
175261 GTTAGAAGGC ATTCCACCA GCTCCATAGT GGTTTATAAA TACCATGGCG ATGTCAGGAA
175321 GCTACCCTAT ATAGTCTAAA AAGGGGAGGA ACGCTTGGTT CTGGGAATTG CCCACATCTT
175381 TCCCGAAAAA CATATGAATA ATCCACTCCT TGTTTAGTAC ATAATCAAGA AATAACTGTA
175441 AGTATCTGTA TTAGTCCATT TTCACACTGC TGATCCAGAC ATACCTGAGA CTGAGTAATT
175501 TATACCAGGA AAAAATGTTT CAGTCTCTTA CAGTCCCACG TGTCTGGGGA GACCTCACAA
175561 CCACAGCAGA AGGCAAGGAG GAGCAAGTCA GGTCTTACAT GGATGGCAGC AGGCAAAGAG
175621 CTTGTGCAGG GAAATTCCTT CCTATAAAAC CATCAGGTCT CATGAAACTT ATTGACTATC
175681 ATGAGAACAG CAGTATAAAT TACTCAGGGA AAGACCTGCC CCCATGATTC AATTACCTCC
175741 CACCAGGTCC CTCCACAAT ATGTGGGAAT TTAAGATGAG AGTTAGGTGG GGACACAGCC
175801 AAACCATATC AGTATCCTTA GTCCAGAAGC TGATGCTCTG CCTGTAGAGT AGCCATTCTT
175861 TTATTCCTTT ACTTCTTGC TTTCACTTTA CTGTGTAGAC TTGCCCCAAA TTCTTTCTCA
175921 CACGAGATCT AAGAACCTTC TCTTAGGGTC TGGGTTGGGA CCCCCTTTCT GGTAACACTA
175981 TCAAAGGATC AGGAAAAGGA AGCTAGTGAA TGCTAAAAAG GAAACAACT ACCATTACCA
176041 ATAATAACAG CAAGACAAA GCAAAACGGA TTGTGACAGC TGTCCCATCT CACACCTGTT
176101 TCCCATTGCA GGAAGGAGGG GCTGTTTCAT GCACAGAGTG GCCAATATTA GAAGCAGAGA
176161 GGGGGTGCAG ATGAGACTTC AGGAATATGT TGACAAAGGC AGGCCTAGGG AGAAATCAAC
176221 CTGAACTATC CCCAAGGAGG AATGCATTAT CTCTAATATG TAAAGTTAGG CTTGATCCTG
176281 TGATTATGGG ATATAGGAGT CCAAAGACTC ACAATGGGAA GTAGGTCACT AGAGTCTCCT
176341 TCAGAAGCTC TGTACTGTGT GTTCCCACTG TGGGCAAGAG TCAGCACTCA GCTATTCCCTA
176401 GAATGCCTTT CCTCAACTCC TTCAGATTTT GCCTCTCAAC TAACCCTATC CTGACCCTT
176461 GTTAGCAAGT GTACCCCTCT CTCCCTCCCA AACATTTTCA AATCTATTTT GTTCCCCTGG
176521 CACTATCAC TGAATATTTT ACTAATTTAT TTTGTTAGT GTTTGCTTCC CTCATGAGAA
176581 TGCAAAGGGA TGGATTTTTT TCAATATTGT TCACTGATGA ATCCAGSTAA CTAGAATATT
176641 TCTAAGCATA GTGATGTGCA TTAATCAAAA GAGTAACCTT CTGAATTGCA CTAAACACAC
176701 ATCACAAGAG GTGTGTGCAC ATATGTGCAT GATGCACGTA GTGTGGTGTG GGTGTTGTGT
176761 GGGGTATGTG GTACTGTGTG TGCTGTGTGT GGTATGTGAT ACATAGTTTG TGTAGTGTG
176821 ATGCATGTGA TGTGGTATGT GTGTGCGTGT CCATACATAT TAGGGGTGGC GGGGATGTTA
176881 ATATGTCAAA TGGTACTAGA AAGTATCAGA ACTCATGGTG CTTACTGGTT TCCCAGAGAG
176941 CTGCTTCTCT CCCACCTGTA GGATATACTG ATGGTTTGGG CAGAGAAGAA ATAAAAAGAA
177001 GGCTGTGACC TACTGGGCTG AGGAAATAAA AACGAAAGTA AAAGAAGAGC TGGGAAAAGA
177061 GAGTGGAGGG GCCAAGGGAA ATTTCCCTT TGGCTTCTGG GGAAACTTTG CTGAAAAATC
177121 AACTCACAAA TTTATTAACA TGTACACAGG GAGAACCATA GAATGATTAT CCACTTCCCA
177181 AGAGGGCTTA AAAGCTTATA TATTATCCTG GCAAAACAGA TTATGGGAGG GGAAGAGAG
177241 AAACCTCTGT GATGGGATTA CTGTTGCGGA TTTTGTCTCC TTCGCTCAGC TAGGTCCGGG
177301 TTTTGTCTC ACAGCCAGGA AGAATTAGGC ATGCAGCCAT CAAAGAATGA GTGGAGTAGA
177361 ATTTATTAAG TGAAAGGAAA GCTCTCAGCA AAGACAAGGG TCCTGAAAGC AGATTCTGG
177421 TTTGCTCTTC ACAGTTGAAT ACTAGGGCTT AAGACTCAAA TTCTTGACAA CTCCACCCTG
177481 TCCTACCAGT GCATGCAGGC CTTTAGACTG AGCTACTCCA TATTGATTAA TTTCTGAAC
177541 TGCGCATGTG TTAAGGAAAG GAATCATCCA CTGCAGGCAT GTTTAGGCAA GCCCCCTGTG
177601 CAAGTTCCCT TATCTGCACA AAACATCCGG TGTAAGCACT TGTGGGGCAG GTCAGAGGTT
177661 CTCTGGGTAC CATTCCCTTA CTGTCTGCCT AAAGCAAGCT GGCCAACTCC TTTCACTACT
177721 AGGGAGAGTA AGTAGATCAG GGAACAGAGA TTAACCTGAA CATTATCTTG TGAAAGTCCG
177781 TTCGGGCATG GTTACATTCT TGGTCTTACA GGAAGGGTAA ATAAAAATAA TTGCTCTTTT
177841 TGGTGGGTCT GGATCTTAGG TAGATAAAGA AACTTTAATT CCACGATGTG TTTTGGTAGG
177901 GATAGTTGGT GGCAGGGATG TCAGAGAGAC TTTGAGGCTT CTTAGTTCA ATATGACCAA
177961 GGGCCATATA TTAGGGTATC AATTTCTGAG CCCCAACAAG AGCTTAGGAG AGATGTGATA
178021 GCATCACAGT GTGAAAGCAA TTTTGTGCT GTTTTATAG ACAGGCTCTT GCACTGTCAC
178081 CCTGGCTGAA GTACAATGGT ACGATCACAG CTCACTGTAA TCTTGAAGT GGTTCAAATG
178141 ATCCTCCCAT CTAAGCATTT CAAAGTGTG GATTACAGG CATGAGCCAC GGTACCCAGC

```

Figure 9 (Page 55 of 74)

SUBSTITUTE SHEET (RULE 26)

144/162

178201 CTGAAACTGC ACCCACTTTC TGATAAACTT TTCAAATGAC TAAAGGGGAG AGAGTAAGCA
178261 CTACTCAGAG GTAGGAAGAA AGGACACAGG ATTATAGGAT TAAACAACA ACCACCAAAA
178321 AAAACCAGAC CGGTGTGGTG GCTCACACCT GTAATCACAG CACTTGGGGA GGCTGAGGTG
178381 GGGGGAGTCA CTGGAGGCCA GGAGTTCGAG ACCAGCCTGG CCAACATAGC AAGACGCTGT
178441 CTCTATTAAA AAAAAAAAAAT ACCTGCCTTG AGCTAATCAG AATCATGGAC CCTGACAAAG
178501 GATGTCCCAA AGTAAGTCTT AGCATTTTTT TTTTTTTTTT GAGACAGTCT CGCTGTGTTG
178561 CCCAGGCTGA AGTTCAGTGG CGTGATCTCG GCTCACTGCA ACAGCTGCCT CCCAGGCTCA
178621 AGCAATTCTC CCTGCCTTCA GCCTCCCAAG TAGCTGGGAT TACAGATGCC CACCACCACG
178681 CCTGGCTAAT TTTTGTTTTT TTTAATAGAG ATGGGGTTTT GCCATGTAA CCAGGCTGGT
178741 CTTGAACCTC TGACCTCAAG TGATCTGCCC ACCTTGGCCC CTCCATAGTG CTGGGATTAC
178801 AGGCGTGAGT CACTGCACCC GGCAAAGTCT TAGCATTCTT TACAAACAGT TTGTACCCGT
178861 ATCTCTAAAA GGGAGTAGT AATTTACCCC CAAAATATGG CTTCTTGATA TAATGAGTAT
178921 TTTGAATGAA AAACCTCTAG AGATCAACAG AACTAAAGA GACTTTTCCC TAGGTACATA
178981 AAAATAGGAT GGCCCCACCA GCGAGAACAA TTGTTCTTTT CTCCCTCCCT GTTATCTCAT
179041 TGTGCATTAT AGGAAAGACC AAGAATGTAA CCACACCTGA ACAGACCCTT TTATAAGATA
179101 ATCAGTCTCT AAGCATCATT TAAATTCCAA GGAGAACTAT TTACAAATTT ATCTGTTCTT
179161 TGATCCAATT AGTCTCTCCT GGTAGTTACA TATTGCCCTT CAACAGAATT CCTCTTCTTC
179221 TGTTTCCCAT AACCTATTTT GCAAGGATCA AGCCCCGTG ACTTCTTCAA CTTCAAGTTG
179281 GCATATAAGC TTCTAAATTC CACTGGGATA TTGGTACTAT GTGCATGAGG AGAACCACAG
179341 AGTAATTAAA TTGTAAAGCC TTTTATCTTA TGAATCTGCC TTTTTTTGTG TTCATTTTTT
179401 AGCAAAACTT CCAAGGGCAA AGGTATAAAA CAAAATATAA ATTCTAAAGC CCCCCAACCA
179461 TCTGAATAGA CTTTCTCTTC AGTCAGCCTT CTTAAATGT AACCTGAAAG ACTGGCTCAG
179521 GCCATTAAGG GAAGTGGGGG TTGAACATGC CTCATTATTC CTCTCTGGCA TTAACATCAA
179581 CACAGCTTTT AAGTCTGATA AGAAACATTT TACAACCTAT TCTCTCTGAA GCCTCTAGC
179641 TAAAAACTTC ATCCCATAGT ACAACTTTGG TCTTCACAAC CTGTTATCAC AACCTAGTGC
179701 TCCTTTCTAT TAATCCCAA TCTTTATACA AACTCAACCA ATTGTATCA CCTCCACCCC
179761 ACTCCTCCGC TGCTTCCAGT TGTCCCGCCT CTCTGGACCA AACCAGTGTA CATTTCTTAA
179821 ACGTATTTGA TTGATGTCCC ATGCCCTCCT AAAATGTATA AAGCCAAGGT GCATCCCAAC
179881 CACCTTGAGC GCTTGTCTC AGGACCTCCT GAGGGCTGTG TCATGGGCCA TGGTCACTCA
179941 AATTTGGCTC AGAATAAATC TCTTCAAATG TTTTACAGAG TTTGGCTCTT GTCATGACAC
180001 AGATGACTGC TCACTGAAG CCGCTCTGG AAGTGAGTGG GGGTTTGTGA AGGATAATTT
180061 TCCCCGGATA GCCCCAGAAG CAGCTAGTAA TAATACACTT AAAGGTAGCT AAAATGCATT
180121 GAACACTTGT TTTGTGCCAG ACCTATGTCA ACATTTGCTT TGTGCCAGGC TTATGCCAGT
180181 ACTCCTGATT TGTTAATACA TTCTAAATAA AAATCTGGA GTTTCAAATA TAATAACTGA
180241 AAAACAGAAA ATAAATAAAA ATATATAATA ACTGAAATAA AAATTTACTA AGGCTGGGGA
180301 TGGTGGCTCA CTCACACCTG TAATCCTGTT ACCGGAAAGG GGTCGGTCCA GATCCAGACC
180361 CCAAGAGAGG GTTCTTGGAT CTCACACAAG AAAGAATTCTG GCGAGTCTG TAAAGTGAAA
180421 GCAAGTTTAT TAAGAAAGTA GAGGAATAAA AGAACGGCTA CTCCATAGGC AGAGCAGCTC
180481 TGAGGGCTGC TGGTCGCCCC TTTTATGTT TATTTCTTGA TTATGTGCTA AACAAAGGGT
180541 GGATAATTCA TGCCTCCATT TTTTAGACCA TATAAAGTAA CTTCTGACG TTGCCATGGC
180601 ATTCGTAAAC TGTCTGGCG CTGGTATGAG CATAGCAGTG AGGACGACCA GAGGTCATC
180661 TCATCGCCAT CTTGGATTTG GTGGGGAGCA GTGAGGATGA CCAGAGGTCA CTCTCATCGC
180721 CATCTTGGAT TTGGTGGGGT TTAGCCAGCT TCTTTACTTT TTTCTTTTTT TTTTTTTTTT
180781 TTTTTTTTTT GCCCAGGCTG GAGTGCAGTG GCACGATCTC AGCTCACTGA AACCTCCAAT
180841 TTCTGAGTTC AAGCGATTCT CGTGCCTCAG CCTCCCAAGT AGCTGGGATT ACAGGCATGT
180901 GCCACCACAC CCAGCTAATT TTTTATATTT TTAATAGAGA CCGGGTTTCG CCATGTTGCC
180961 TACGCTGATC TCCAACCTCT GCGCTCAAG CATCCAGCCA CCTTAGCCTC CCAAAGTGCT
181021 GGGCTTATAG GTGTGAGCCA CCCCACCTGG CCTAGCCGGC TTCTTTACTG CAACCTGTTT
181081 TATCAGCAAG GTCTTTATGA CCTGTATTTT GTGCCCCACTG CCTGCCTCAT CCTGTGGCTT
181141 ACAATGCCTA ACTTACAGGG AATGCAGCCC AGCAGGACTC AGCCTTATT CACCCAGCTC
181201 CTATTCAGA TGGAGTCTTT CTTGTTCAA TACCTCTGAC AAGCCCAACA CTTTGGGAGG
181261 ATGACACAGG AGGATTGCTT TAGCCTAGGA GCTCAAGACC AGCCTGGGCA ACACAGTGAG
181321 ACCCATCTC TAAAAAATAA AAATACAAA AAATTAGCCA GGCATGATG TGTGTGCCTG
181381 TAGTCCCTGC TACTCAGGAG GCTGAAGTGG GAAGATGGCT TCAGCCAGG AATTCAAGGC

Figure 9 (Page 56 of 74)

SUBSTITUTE SHEET (RULE 26)

145/162

181441	TGCATTGTCA	GAGGCATTTG	AACCAGAATG	ACTCTATCTT	GAATAGGGGC	TGGATAAAAT
181501	AAGGCTGAGA	CCTGCTAGGC	TGCATTTCCA	GTATGTTAGG	CATTCTTAGT	CACAGGATGA
181561	GATAGGAAGT	CAGCACAAGG	TACACATCAC	AAAGACCTTG	CTGATAAAAT	AGGTTGTGGT
181621	AAAGAAGTTG	GCCAAAACCC	ATCAAAACCA	ACATGGCCAC	CAAAGGGACC	TCTGGTTGTC
181681	TTCAGTGCTC	ATTATATGTT	AATTATAATG	TATTAACATG	CTAAAAGACA	CTCCTACCAG
181741	CATCATGACA	GCTTACAAAT	ACTGCGGCAA	TATCTGGACT	TTACCTTATA	TGGTCTAAAA
181801	GGTGGAGGAA	CCCTCAATTT	TGGGAATTGT	CCACCCCTTT	TTTGAATGTC	TCATGAATAA
181861	TCCACCCCTT	GTTTAGCACA	TAATCCAGAA	ATAACTATAA	GTATGCTTAT	TTGAGCAGAC
181921	CACGCTGCTG	TTCTGCCTAC	AGAGTAGCCA	TTCTTTTATT	TCCTTACTTT	CTTAATAAAC
181981	CTGCTTTTAC	TTTACTGTAT	GGACTTGCCC	TAAATTCTTT	CTTGTGTGAG	ATCCAAGAAG
182041	CCTCTCTTGG	GGTCTGGATC	AAGACCCCTT	TCTGGTAACA	TCTTTCTGGT	GACCACGAAG
182101	GGACAATACT	GAGGAGACTC	TGAAGCCAAA	GGAAACAGAC	TACAGCACCA	ACTGGCTGAC
182161	TTTGGGTAAG	TGGTGGAGTC	CCCGGGTAAA	GGATAGGATT	GGGTTAGAGG	TGCAACTTAG
182221	GGGAGATAGG	GTCCTCCTA	AGACAGAGAG	CGTTTCAGTC	CGCTCTTAAT	AAAGGGCAAG
182281	AATGCTTGAC	CGAACTTGGG	TTTGAGACCC	AACCTTAGGAA	GGCTACAGTC	CTTAAGATTT
182341	AAGGGGTTAG	AGGCCCCCTC	CAGTAAAGTC	TCTCTTGGTT	AAAAACGGAT	TTAGCATTAG
182401	GGGATGTTAA	CTGCTATTCT	GTTTGTATTA	ATCTTCCCTG	TGCTCTTTGC	TGACAGCTAT
182461	GGGTGACAGG	ATTAGGCATG	TACAGGATCA	CGGGACATTG	GGAACTTTTC	TTCTCTCCAA
182521	AAGGGGAAGC	TTGACAGCTG	ATAGGACTGT	TGGAAGAGAT	CCCTTTGCTA	TGACCAAGCAG
182581	CCGCTGAAC	TTTTGATTCA	GTGTTGCTGC	AATGGGTGGG	TCTTTCTCTG	GCCTCTGTGA
182641	ACTCCTCACC	TTCCCCACCT	CACCACAGGC	AATGCTTTTC	TCCCTTTCTC	TCTTTTCTCT
182701	TTTCTGTCTT	TTCTGTTACT	TGAGACAACC	ATCTTGCCCA	GAGACCATAT	GTTGAAACTC
182761	CTGGTCAGAA	GTTTGATTAA	AGATGAAAGG	GCCTATCTGG	GGGCAAGTTT	GAGCCTTCCC
182821	AGTTAGATAT	TGGGTGCTAA	GTTGGAGTGGC	CAATGTCTAT	GTTTGTGCAC	ATGTATATTG
182881	CTCTGGCTGA	AATGGAAAAC	GTTAATTTGG	TTACTTTATG	TGGCCATTGG	GCAGCATCTT
182941	ACAAAAGTGA	GAGACATTTA	TTTGCCTGTG	GTTCCATGAA	ACAGAAAAAA	GTTGGTTTTT
183001	CTTTGTGTCG	TAGCTTGGAC	CCAAGGGCTT	TGCAGTGAGC	AAGGTTGCTA	CGCCTGCTCA
183061	GTGAAAGAGA	ACCCAGAAAAC	CTGGCATGCC	AGCAAAAGGG	TAAAGATTTT	TTACCAGTCA
183121	GGCTTCTGGC	CTCTCTCTCT	TAGTGAAAAC	TGAATGAATG	TGAAAAATCA	CTGTTTATCA
183181	CCTCTGTAAA	GTTTGTGATTA	ATGGGAACAA	GGATTTGTGG	GGCTAGTCTT	AAGCTGTAAT
183241	GAATCTGGTA	TACTTTGTGA	TATCAATTTG	TCTTTCTGTA	TTACTCTGTC	ATAAAAGAGGA
183301	ATATGGTAGG	ATAGAACATG	GGCTTAGGAC	TCCATAAGCC	TGCTGTTCAA	GCCAGCCCAG
183361	TAAACTGGTC	CGTTGCAAAG	TTTATTACAG	GTCCCTGGAA	AAAAAAAAAA	TTAAAACTG
183421	GATGAAGTTT	CCTTCTCATC	TTGTTTTATG	TCCTTTGGAG	CTTCACCTTG	TAACCACGTG
183481	CGGGTACTTT	CTCTTGGTCT	CTGCCATCCA	GGGAACAGGA	ATTTTGGGGT	TTATGTAATA
183541	GTTAACTCTA	AAAATTATCT	CAAGCCATTG	CAAGCTCAAA	ATTGGCTGCT	CTGGACCCCT
183601	TCTGGGAAGG	GCAATGGAAA	CTAACCAGTG	TTGTAGCTCA	GCAGCTAAGG	ATTGTGCATT
183661	TTATAATGGC	GGCCAAGGTT	CAATCCTGGC	TTAGGGAATG	AGTACTTTCT	GATTGATATC
183721	TGTGTGACCT	TTACCATTTG	TTGATTCTGT	TCTCTTCCCC	TCCACACACT	GTCTTGAGTT
183781	TTCTCTCTCT	TGAGAACCTG	GGAGATTATC	TTTGGTAAAG	TTCAAAAGCC	AGAAATAATG
183841	GCCGTGTGGG	ATGGCTAAAG	TTGAGTAATA	AGAAACTTAA	AAGGACTCCT	TTTTTTTTTG
183901	CTTTAGAGTG	CTATGGTTTA	TGGTTAAAAG	CTTAATTAAA	AGTGGATATT	CAATCTCTAA
183961	AAGCCTGGGA	CTCCTTGGGA	AAAGCAGAGG	AGGCACCACA	GACCCCATTT	TGGGAAAACC
184021	TCTGTTTTCC	TCATGAAACC	CCAGGAAC TG	GAAGTGGATA	GATCCTTCGC	AAAATCTAAG
184081	GCTCTGTTTG	GCTTTGCATT	ATGTTATCTG	ATGTTTTTGA	CTTTTGGGGG	TATCAGAAAT
184141	TACTTTGTCAT	TATGAGGGAG	ATCTGGTGTG	TAATAACCAG	GTAGGAAATA	TACTTCTGGG
184201	GATAGCTAAA	GGCAAATATA	GGTGAATACT	TGGCTATTTG	CACCTTTTGA	TCACAAGAAG
184261	CATTCTCTTG	ACTACCTAGA	AGGTATGGAA	ATGTCTCCAT	CCCCACCGAG	AGATAAGATT
184321	CCCAGGGGAG	ATGGCTGATC	CCCCAAAAGA	GGGCTGATTC	CCTCTTTTGG	GATCCAGGAT
184381	CTGGTATAAA	AATGGGACCC	TGGCCAGGCA	CAGTGGCTCA	CGCCTGTAAT	CTCAACACTT
184441	TGGGAAGCCT	CAGAGTTATG	AATGTCTCAC	CATACTGACA	CTTTGTGACT	GAGCTCCTCT
184501	CTACCCTGGA	CACAAGAGAC	CCTAATAATT	AGACAGGAAT	ATCATTGCCC	CTATTTAGTC
184561	TGAAGAAGTT	ATAGAAGATG	GATCTTTATC	CCACTGCAAT	CCTTAGGATT	AAGGGTCCCC
184621	TGGTAAAAGG	GAGTGGGAAA	ATATGTGAGA	GGCATTTGAA	TCAGAGTGAC	TCCATCTTGA

Figure 9 (Page 57 of 74)

SUBSTITUTE SHEET (RULE 26)

146/162

184681 ATAGGGGCTG GGTAAAATAA GGCTGAGGCC TGCTGGGTTA GGTTAGGCAT TCTAACCAGG
184741 AGTTTAGTCA CAGGATGAGA TAGAAGGTTG CACAAGGTAC CCGTCACAAA GACCTTGCTG
184801 ATAAAATAGG TAACGGTAAA GAAGCCAGCT AAAGCCCACC AAAACCAACA TGGCCACAAA
184861 AGTGACCTCT TGTCATCCTC ACTGCTCATA TACACTAATT ATACTGCATT AGCATGCTAC
184921 AAGACACTCC CACCAGTGCC ACGACAGTTT ACAAATACCA TGACAACATC TGGACGTTAC
184981 CTTATATGGT CTAAAACGGG GAAGAACCCT TAGTTCTGGG AATTGTCCAC CTCTTTCCTG
185041 AAAAAATTCTT GAATAATCCA TTAGTTTAGC ACATAATCCA GAAATAACTA TACGTCTGCT
185101 TATTTGAGCA GTCCATACTG CTGCTCTGCC TATGGAGTAG CCATTCTTTT CTTTTATTTT
185161 TATTTTTTAG ATAAAGACTC GCTCTGTCAC TCAGGCTGGA GTCTGGAGTG CAGTGACGTG
185221 TTTTGGCTCA CTGCAACCTT CACCTCCCGG GTTCAAGCAA TTCTCCTGCC TCAGCCTCCC
185281 AACTAGCTGG GACCACAGGT GGGTGCCACC ATGCCTGGCT AATTTTTGTA TTATTAGTAG
185341 AGATGGGGTT TCGCCATGTT GGCCAGGCTG GTCTCGAACT CCTGGCCTCA AGCGATCCAC
185401 TTGCCTTGGC CTCCCAAAGT GCTAAGATTA CAGGCATTAC CCACTATGCA TGACCCATTTC
185461 TTTTATTTCT TAACTTTTTT TTGTTTTTTT GAGACAGAGT CTCACTCTGT CACCCAGGCT
185521 AGAGGCTGGA GTGCAGTGGT GCGATCTTGG TTCACTGCAA CCTCTGCCTC CTGGGTTCAA
185581 GCGATTCTTC TGCCTCAGTC TCCTGAGGAG CTGGGACTAC AGACATGTGC CACTACACCC
185641 AGCTAATTTT GTATTTTTAG TAGAGACAGT GTCTTGCCAT GTTTGTCTCGA CTTGTCTCGA
185701 ACTCCTAACC TCAAGTGGTC TGCCTGCCTC AGCCTCCCAA AGTGCTGTGA TTACAGGCAT
185761 AAATCACTGC GCTCGGCCCT TCTTTACTTT CTTAATAAAC TTGTTTTTAC TTTACTGTAT
185821 GGACTAGCCC CAAATTCCTT CTTGTGTGAG TTCCAATAAC CCTTTTGTGT GTGAAAGAAT
185881 TTATGGCTGC TGTTCAAGCT GGAGCAAGCT GGAGCTCATG CTGCTGCTCA GACTGGAGCA
185941 TCGGTGATCT GTGATCCCAG TAAGAGGATC ATGGTCACTC CAGCCTGAAC GACAGCATGA
186001 TATCTCATCT GTAAGAAAAA AAAAATTACT AGAGGGCTTT AACAGCAAAT TTGAGCAGCA
186061 AAAAGAAAGTA ATCAGTGAAC TCAAAGATAG GTCAATTGAA ATGATCTACT CTGAAAAACA
186121 GAAAGAAGAC AGAATGAAGA AAAAGAAATA GAGCCTTAGA GACAGGGGAT ACCATCAAGC
186181 ATACTAATAT ATGCATAATG GGACTCCTAG AAGGAGAAAA GTGAGAGGAC AGGGAGAGAG
186241 AATGTTTGA GAAATAATTT CTCAAAGCTT CCCATGTTTG GCAAAAAAAC ATTAACCTTG
186301 ATACATATTT TAGGAGCTCA ATGAATTCCA AGTAGGATAC ACTCAAAGAG ATCCATACCT
186361 AGACACATCA TAATCAGATT ATCAAAGAT GAAGAAGATG AATCTTGAGA GCAGAAAGAA
186421 AGGAACAATT CATCACATAC AAATAGTACT CAAAAGATGT CTGGAGTAGG TATACTAATA
186481 TCAGACAAAA TAACTTTTAA GATAAGCATT GTTATAATAA ATAAAGAAAG GTATTTTGTG
186541 ATGATAAAAG TGTCATTCA TCAAGAAAAA ATAACATTAT AACATACAT GCACCTAACA
186601 ACAGAGCCCT AATATTCATG AAACAAAAT GACAGAATTG AAGGGAGAAA TAGAAAATTC
186661 GACAATAATA GTTGAGACA TCAATACCTC ACTAGTTAGA CAAGATCAAC AAAAAATAG
186721 AAGACTTAAC ACTTGAAAAC ACCTAACCTG ACCCTAACAT AAATCTATAG GTCACTACAC
186781 CCCAAAACAG CAGAATAAAC ATCCTTCTGA AGCTCACATG AAACATTTT CAGGATAGAC
186841 TGTATATTAC TTCATGAAAT AAGTCTCAAT AAATGTAAAA GGACTATAAT AATAGAGTAT
186901 ATATTCTCTG ACCAAAGTGG AATGAAGATA GAAATCAATA ACTAGGCTGG GCGTGATGGC
186961 TCACGCCTGT AATCCCAGCA CTTTGGGAGG CCAAGGCGGA CAGATCACGA GGTGAGGAGT
187021 TTGAGACCAG CCTGACCAAC ATGGTGAAAC CCTGTCTCTA CTAACAAAAT ACAAATAA
187081 GCCAGGCCTG GTGGCATCTG CCTGTAGTCC CAGCTACTCG GGACACTGAG GCAGGAGAAT
187141 CACTTGAACC CAGGAGGCAG AGATTGCAGT GAGCTGAGAT CGCGCCACTG CATTCAGGCC
187201 TGGGAGACAG AGCGAGACTC CATCTCAAAA TTAACAAAAA AAAAGAACT AGAAAAATAA
187261 GAACAAATCA AACCCAAAGC AAGCAAGAGG AAAATGAAAA ATTTCAAAGC AGCCAAGAAC
187321 AAAAGGCACA TTATGTACAG AAGAACAAGT GTATAGATCA CATATTTCTC ATAGACACAA
187381 TATAAGCAAA AAGACAGTGG AGCAAAATTT TTTAGATTAA TGAAAGACCT ACAATTCTGT
187441 ACCAAGCAAA AAAACTCCCC CCAAATGAGG GTGAAATAAG ACAATTTAAT ACAGAGAAAA
187501 GAGGAAGGAA TTTATCTAGT CATATGTGAG AGTTTTATGA TACATTTTGT ACTGTATATG
187561 TGGATGTTTT CTATTTTCAAT TAAAAATCA ACCGTGCAAT TAAATGGTAG ATTGTCTTGC
187621 TTCTTTTTGA TTGACACAGT CATTAACATA AATATTGTAG TATTTTTTTA TCTCCCTGCC
187681 TAAAGGCAAT AAACATCTAA TCAGCAGACT AGAACAATAA AAAATATTTT TTAAGGTCC
187741 TTTAGGCAGA ATGATAAAAG TCCCTTAGCC ATATTGAAAT TCCTATTAT ACAAAGGAAT
187801 AAACAGTACT AGAAATTGTA ACTATGTGAG TAAACAGATA ATATTTTTTC TCCATAAAAT
187861 GTGGTTGACT ATTTTCACAA AAATAGTTAA CAATGTAATG TGTGATTAT AGCATTTAAA

Figure 9 (Page 58 of 74)

SUBSTITUTE SHEET (RULE 26)

147/162

```

187921 AGTAAACAG GCCGGGCACA AAGGTTTCGT CCTGTAATCC CAGCACTTTT GGAGGCCGAG
187981 GCGTGCAGAT CACTTGAGGA CAGGAGTTCA AGACCAGCCT GGCTAACATG GCAAAAACCC
188041 ATCTCTACTA AAAATACAAA AATTAACCAG GCGTGGTGGT GCACGCCTGT AATCCCAGCT
188101 ACTCTGGAGG CTGAGGCACA AGAATCACTT GAATCCAGGA GGTGGAGGTT GCAGTGAGGC
188161 AAAATTATAC CACTGTGCTC CAGCCTAGGC AACAGAGCTA GACTCTGTCA CACACACACA
188221 CACACACAAA AGAAAAGTGT ATGACAACAA CAGTGCAAAA GAAGCGGAAA TGAAAATAAT
188281 GTTATTTTAT ATAAGTGGTA TACTTTTAGA TGAACACGA TAAATTAATG ATGTATACTA
188341 TAAACTCTAA GGCAACCACT GAAATAATGA AACGAAGAAT TATGGCTAAC AAGCCACAAA
188401 AAGAAATAAA ATAGAATGAG AAAAAATATT TAAGTTGTTC AACAGATGGG AAAAAAAGA
188461 GGAAAAGAG AACAAAGAAC AGATGGGACA AATGGGAAAG TAATAGCAAG ATGATAGACT
188521 TAACTCTACC CATATAGATT ATCACACTTA AGGTAAATGA TCTAAATACT CTAATACAAA
188581 AGCAGAGGTT GTCAGATTGA ATTAACAAAA CAGACAACAA CAAAAAAG CAAAAAAGA
188641 GCCACAACAT GCTGCCTACA AAAAATTCAC TTAAATATAA AGACACAAAT AGTCTAGAAC
188701 ACCATCACTT TTAACCTTAT TTAACCAAC CTCCTAAGT ATCCCTATTT ATTTATTTAT
188761 TTATTTATTT ATTTATTTAT TTATTTTGA GACAGAGTCT GACTCTGTTG CCCAGGCTGG
188821 AGTGCAGTGG CACCATCTAG GCTCACTGCA GCCTCTACCT CTCGGGTTCA AGCGATTCTC
188881 CTGCCCTCAGG CCTCCCAAGT AGCTGGGACT ATAGCACATG CCACCATGCC CAGCTAATTA
188941 TTATATTTT AGTAGAGACG GGGTTTGGCC ATGTAGGCCA GGTGGTCTC AAACGCCTGA
189001 CCTCAGCCTC CCAAAGTGCT GGGATTACAG GCGTGAGCCA CAGCACCAG CTCCTCTTCA
189061 TTTATCTTG CTACGCTTCC TCCAATCCAT TTTGTGCATT TGATGATTTT GCCAGTAAC
189121 TCTTTATTTT TCTGGTAAAA TTACTTATGG GTCAGTGGG ACTGGGATGT TCTTTCTTCT
189181 AGAGGGGGTT TGTGTCTGCT TTTGCCAGGA AGCTGGGGTA CCACCACTCA AGTATTACTT
189241 TAAACTCAAT TCATGAATTG AGACTTTTTT TTTTTTTTTT TTTTTTACGC AGAGTCTTAC
189301 TCTGTCAACC AGGCTGGAGT GCAGCGGTGT GAACATGGCT CACTGCAGCC TCAACCTACT
189361 GAGCTCAAGC AATCCTTCTG CCTCACCATT CTGTATAGCT AGGACTACAG GTGTGTGCCA
189421 CCATCCCTGA CTAATTTTTT AAATGTTTTT TTAGAGATG GGGCTCACTT TGTTGCCAG
189481 GCCGCTCTCG AGCTCCTGGG CCAAGTGATG CTTCCACCT TGGTCTCCCA AAGTGTGGG
189541 GTTACAGGCA TGAGCCTCTG TGGCTAGCCA AGACTTTTTA TTTTTTAGCC TAAATGTGTA
189601 TAAAGTTGG CTTGTGGTTA CAACTTATCA GGATTGATGA TCTCTCTCTC TCTCTCTCTC
189661 TCTGTCTCTC CCCACCTCTC TCACATCCCT TGCTCTGCTG AGAAGCAGAG CAAACATTCT
189721 AGCAGTTTCC AGAGAGTAGG ATGGGATTAC TTCTAGTTTA CTTTATCAT CTTTGGGAT
189781 CGCAGTATTA CTGGGAGAAC ACAAGTATCT CTTATTAGAC ATACCACCTT TGTAGAATCT
189841 GGACTTTCAT TTTAGACTTT ATTTGTTTTT TACTATAAGC AATTTAAGTT ACAGATCTCT
189901 CTACACACTG TTTAAGTTGC ATCCCATGAA TTTTGATGTG CTTTATTGTC ATTATTATAT
189961 AGTACAATGT ATTTTGTAAT TTTTGTGAT TTGTTTGGAG AGATTGATTA ATTAGAATGA
190021 TGTTTAAATT CCAAATATGT GTGTTTTTTT CCTACATTTT TTATTTTAT TGATTTCAAA
190081 TTTATTTCTA CTGTAGTCAG ATTTAATAAT TCATTTTATT TTATTATTTT CATTTTTTTA
190141 GAGACAGGGC CTTTCTGTGT TGCCCAGGTT TGTCCCAAAC TCCTAGTCCC AAGCAGTTCT
190201 CCTGCCTCAG CCACCCAAAG TGCTGGGATT ATAGGCACGA GCCACCCGTG CACAACCAAC
190261 AATTCATTTA AAAAGTGGGC AAGTGAACG AACAGACATT TCTCAAAGA AGGCATACAA
190321 TTGGCCAAAC AATATATGAA AGAATGCTCA ACATCACTGT ATTAGTCTGT TTTTCATGCTG
190381 CTAATAAAGA CTTAACCTGA GACTGGGGAA TTTACAAGAG AAAGAGGTTT AATGGACTTA
190441 CAGTTCCACA TGGCTGGAGA GATCTCACAA TCATGGTGGA AGGCAAGGAG GAGCAAGTCA
190501 CATCTTACAT GGATGGCAGC AGGCAAAGAG AGAGCTTGTG CAGGGAACT CCCGTTTTTA
190561 AAACCATCAG ATCTCGTGAG ACTCATTCAC TATCATAAGA ACAGCATAGG AAAGACCCGG
190621 CCCATAATTC AGTCACCTCC CACTGGGTTT CTCCCAGGAC ACATGGGAAT TGTGGGAGTT
190681 ACAATTCAG ATGAGATTTG GGTAGGGACA CAGCCAAACC ATATAAATAA CTAATCATCA
190741 GGGAAATGCA AATCAAAACC ACAATAAGGT ATCATCTCAC CCCAGTTAGA ATGGCTATTG
190801 TCAAAAAAAC AAAAAATAAC AAATGCTGGT GAGGATGTAC AGAAGAGGGG ACTCTTATAT
190861 CCTACTGGTG GAAATGTCAA TTAGCATAGC CATTATGCAA AATAGTATGG AAGTGAGGTA
190921 GGTTACATAG GGTGGTCACA GCCTCCCTTG AAAGGAAACA AGAACTTGT CAAATTGATG
190981 GAGAGAACAA ATCTCTGAC ATTACACAAA CTGCATCTGG GGCTAGTGGT TAGAATATCC
191041 TCAGTCAAGG AGGTAGAAGA GCAGGAGGGA AAATCCCTAA GTTCGTGCAA GTGCAGAAAC
191101 CCACAAGCTG TGTTCTCAGG TTGACATATA CTCATTTTAA TAGTAAGAAA CACACCCTTG

```

Figure 9 (Page 59 of 74)

148/162

191161	GGTAGAGAAT	TAAAAATGCTA	ATAATACATG	TGATGTATGT	ACTAGCGTGT	ATGGCAATAT
191221	TGCATGCACA	TTCAAGAGAC	CACCCAAAAC	ATATTTAACA	ACAATGCCCA	TTCCCACCCC
191281	CTCATGGATA	ATCACGTAGG	ACTCCCATAA	CGGGAGTTTC	TTCAAGTGTCA	ATTGGTGCTG
191341	AAGTAGCCGA	CCCTGACTCT	GCTATCAGCG	TGTACTTTCA	CCTTGCAATA	AACCTCCTTG
191401	CCTACTTTTA	CTTTGGACTG	GCTTTCAAAT	TCTTTTGTGC	AGGGAATTCA	AGAATCTGAA
191461	CCAGCCCCT	GACAACAGAG	GTTTCTCAGA	AACCTAAAAA	TAGATCTACC	AGATGAGGCT
191521	GAAAATCTGC	TACTGGCTAT	TTATCCAAAG	GGAAGGAAAT	CAGTATACAA	AGAGACACCT
191581	ACATCCCCAT	GTTTATTGCG	TCACTCTTCA	CAAGAGCTGA	TATATAGAGT	CAACCCTAAA
191641	TGTTCAATTAA	CAGACAAATG	GATAGAAAAT	GTGGCATATA	TACACAATGA	AATACTATTT
191701	GGCCATGAGA	AGAATGCAAT	CTTGTCATTT	GTGGCAACGT	AGATGAAACT	GGAGAACATT
191761	ATGTTAAGTA	AGATAAGCTA	GGATTGGAAA	GATAAATACT	ACATGTTATC	ACTCATATGT
191821	GAAAGTAGAG	AAAAATTTTT	AGCTCATGGA	TTTAGAGAAC	AGAACTGTGG	TACCCGGAAG
191881	CTGGGAAGGG	TAGCAAGGAG	GGGAGGATAG	GGAGAGGTTG	GTTAATGGTG	ACAAAATTAC
191941	AGCTAGATTG	TAGAAATGAG	TTCCGGTGTT	CTGCACCATT	GTAGGGTGCA	TATGGTTAAC
192001	TCTCATTTAT	TGTATATTTT	CAAAAAGCTA	GAAAAGAATT	TTGAATACTC	ACAACAAAAT
192061	AAATGATAAA	TGTTTAAGGT	GATGGATATA	CTAATTACTC	TGATTTGATT	ATTACACATT
192121	GTGTACACAT	ATAAAAATAT	CACTCTTTAT	CCCGTATATA	TGTACAGTTA	TTATATGTCA
192181	ACTAAAAATA	AAAGAAAAAA	AGAATATGAT	CTATCATGAT	GTATATATCA	TGTGTACTTG
192241	AGCAAAATGT	GCATGCAGAT	ATTGTGTATA	ATGTTCTATA	AATCAATTAG	CTCAAGATAA
192301	TAGATAGGAT	TGTTTCAGATC	TTCTGTGTCT	TTACTGATAT	TTTGTCTAGT	TATTGCATCA
192361	TTACCAAAAA	AAGGGTGTTA	AACCTCCTCA	ATGTGATTGT	AGAATTGTCT	ATTTTGTCTT
192421	TTCTTTTCCA	TTTTTACTTT	ATGTATTTTG	AAACTCTGTT	ATGACATTTT	GCTATCTATT
192481	TTAAAACTTC	GTTATGTATT	TGAAACTCT	GTTGTTAGAA	TCATACATTT	ATGATTATTA
192541	TGTTTTCTTG	ATGAAATGAC	CCTTTTCTAT	TGTCGTTGTT	TTTGTTTTTT	CTGAAATGGA
192601	GTCTCACTCT	GTTGCCCGAG	CTGGAGTACA	GTGGCACAAAT	CTTGGTTCAC	TGCAACCTCC
192661	ACCTCCTGGG	TTCAAGCGAG	TCTCCTGACT	CAGCCTCCAA	GTAGCTGGGA	TTACAGGCAT
192721	GTGCCAGCAT	GCCAAACTAA	TTTTGTATTT	TTATTAGAGA	CAGAGTTTCA	CCACGTTGGC
192781	CAGGCTGGTC	TCGAACCTCT	GACCTCAGGT	GATCCGCCCA	CCTCGGCATT	TTTATTTTAT
192841	TTTATTTTTT	TGAGACAGAG	TCTCACTCTG	TCACCCAGGG	TAGAATGCGG	TGGTGTGATC
192901	TTGGCTCACT	GCAACCTCCG	CCTCCTGGGT	TCAAGCAAAT	CCCATGCCTC	AGCCTCCCGA
192961	GTAGCTGGGA	TTACAGGCAC	ATGCCACCAT	GACTGGCTAA	TTTTTGTATT	TTTAGTAGAG
193021	ATGGGGTTTT	TCTATGTTGG	CCAGGCTGGC	AACTGACTCC	TTTAACAATA	CAAAATATCA
193081	CTCTGTCTCT	GGTAACACTC	TCTGTCTTAA	ACTCTATTTT	AGCTGTTATT	ATTATAGCCA
193141	TTTTAGTCTT	TTTATGCTTT	CTGTTTGCAAT	AGTGTATATA	TTTTAATATG	TTTATTCTCA
193201	AGTTATCTGT	GTTTTTATAT	TTAAGATGTT	TCTCTTCTAG	CCAACGTGTT	TGGTTCCTGC
193261	ATTTTTAAGT	CGATTCTAAC	AATCTTTGCC	TTTCAATTGA	AATATTTTACA	CCATTAAACAT
193321	CTAACATTAA	CATTTATTTT	TCTTTCCACA	GTACACTGGC	TAGCATCTCC	CATATAATAT
193381	TGAACATAAA	GTGTGATAAC	TGACATCTTT	ATTTCAATCC	TACTCTGAGT	GGAAAGGGCA
193441	GGGGTGGAGA	AAGCATTCAA	CAATTTGCCA	TAATTATAAT	TCTTTTTGTT	ACACTGTTTT
193501	CTTCTGCATT	AAAAAATATC	ATTACATTTT	GCATGAATTA	TTAGGAGAAA	ATATTTTCCA
193561	ATTTTCCTGG	AAAATGCCAT	AACCACGTCT	CTCAATTTTG	TTTCCATCTT	TCTTCCACAT
193621	TTTACATAAC	CTACATAAGA	GACACATTAT	CAAGTATATT	TTACATGGCT	TCTCAGTGTC
193681	TTCTCTGTCT	GCTAACAGGT	TTACCAAGAG	ATGGCACTCT	TGTATTTCTG	GTGGCTATGT
193741	CCATATCGTT	TTGCCTTTAA	GACAGCGTAA	CTACTTCTTT	CACCAGTATT	AAAGACATGT
193801	ACATTTGATC	TGGTTCCTGT	GGATGATTTT	AAATGACTCA	AGCTAATTAAT	CCTAATTTTA
193861	CCTAAACACT	CCATTATTTT	AAAATGTATT	CCTTTATGCC	CACAATAAAC	ATTTATTGAC
193921	ATTAGGCTGG	ACATTAGGCT	TCTCTATGGC	AGACATTAGG	CTGGACCCCTA	GCCATATATC
193981	TATTGAGGGA	AAAAAAATTA	TTTTCTATAT	AAGTTTCCAG	AAAGCCAGA	TGTGTTTTAA
194041	AAACAAAACA	AAACATTACA	TTCTAAATGC	TGTAACAAGA	TAAGAAAAAG	TGTTGAGGCT
194101	GAGAGAAGAA	CAAAGCAGCA	AGCAACTCCT	GGAAGGACCA	CTGCTGCAGA	GGTAATAACT
194161	GGTGAACCAT	GTTTTGGAGA	AGGAAAAGGT	CACCAAGAGA	AGGAGGGGGT	CCAGGGTGTT
194221	CAGAAAGATT	GCATGCATAA	AGATCAAGGG	TAATAAAAAA	AATTCCGTAT	TATGTAAATG
194281	TGAAGTTCCA	GGACCATGAG	CTTGGAGAGC	ATGAAGTACA	GGAGGAGGGT	TGGTTTCAAA
194341	TAAATCTGGG	AATGAAACAG	TGAAGCCTCT	GGCAGAACTC	ACATCTCTTT	CCTCCCTCT

Figure 9 (Page 60 of 74)

SUBSTITUTE SHEET (RULE 26)

149/162

194401 TCCTTGACACA TTCCCTTTAT GGAGTAATTG CAGGGATGGG AAAAGTTCAA AACCACCACT
 194461 GAGCCTAGGA AGTGCTAGGG TAAAGTGGAG AATGAACCTG CGTGATTTGC TCATCCTAAA
 194521 CTAGGTTCTT CTAGGAGAGC CTTTCCCAT AAAATCTGCC CTCCTCGAAG GGGCCCAGAC
 194581 AGCCTAAGCT CACCTCCCAA AGACCCCTTA CTTGCTGACT GAATCTGATT CCACCCAGAC
 194641 ATGGCCTAAA ACCCTTCCAT AACTCTATAG CCAAATTCAA TTTTAGACAG GCCTCATACC
 194701 AACCTTTCTT CCTCTAAGTC TGCCACCCTA GGCAATTCTC AACATTCTCT ACACACTTTG
 194761 GGGCCATAGA CGTGCTACCA AGTCTCCAGA CCTAGACCTG ATGGAGCAGT GCTGTAATGA
 194821 GACGACCACT GGCCTTTGAA CCAGACCCTT CTCTGTGGCT CCTATGCATC TCCAACCTGT
 194881 TTTGAGCACT GCTGCCAAGA CATCTTTGGC ACTTTGTTGT GAAGTTTTAA AACTGAACTA
 194941 ATCTACAAAA CACCTAACCT TTAAAAATTC ATTGTCATTT CATATCATGA AAGATAAAGA
 195001 AAGGCCAGGA AACTGTTCCA GGTAAATAGA GACTAAAGAG ATAGCAACCA AATGCAATTT
 195061 GTGATCCTGG ATTGAGGGGA AAAAGTGTG TCAGAGACAT GATTGGGACA GCTGGTAAAA
 195121 TTTGAATTTG AATTTAAAGA TAAAGTATTG AGTAATATAG GAAGATGATT ATCTGCAACT
 195181 TTCAAATGTT TCAGTAAGTA TATATATATA TAAAGAGATA TAAAGACATA TAAATAAATA
 195241 GATGGATAGG TAGAGAAAAA GCAAATGTAT AATATTAACA ATCTAGGTAA AAAGTATATG
 195301 AGTGTCTTTT GTACTGTTTT TCTGATTTTT CTATATGTTT GAAATCATTT TAAAAAAGA
 195361 AGGTTTTTGG GGTTTTTTTG TTTGTTTTTT GTTTTTTAGAG ACAGCATCTT ATTCTGTCAC
 195421 CCAGGCTGTA GCTCAGTGGC CCAATCATTG CTCACTGCAG CCTCAACTTC CTGGGCTCCA
 195481 GTAATTCCCC CTACCTCAGG CTCATGAGTA GCTGGTACTT CAGGTGTGCA CCACTGCACT
 195541 CAGCTAATTT TTATTTTTTA AATTTTTGTA GAGATGGCAT GTTGCTATGT CACCCAGGCT
 195601 AGTCTCAAAC TCCTGCCCCC AAGTGATCCT CCCACTTTGG CCTCCCCAAG TGCTAGAATT
 195661 ATAGGCATGA GCCACTGCAC CCAGCCCCAA ATAAAAAAGT ATTTTATTTT AATTAACATA
 195721 TTAATTTTGA GTCAGAGTTT CACCCCTGTC ACCCAGGCTG GAGTGCATG GCATCATGTT
 195781 GGCTCACTGC AAACCTGCTC TCCTGTGTTT AAGCGATTCT CTTGCCTCAG ACTCCTGAGT
 195841 AGCTGAGATT ACAGGTGCCT GCCACCATGC CCAGCTAATT TTTATATTTT TAGTAGAGAC
 195901 GGGGTTTCAG CATGTTGGTC AAGCTTGTCT CAAACTCCTG ACCTCAGGTG ATCCACCCAC
 195961 CTCGGCCTCC GAAAGTGTTG ATGAGCCACC ACACCCGGTC TAAAAAGTAT TTTAAACCA
 196021 CAGTCCCACT CTACCTTGTC CTACACTACC AGGGGCTAGG ATCACCCCAT GTCTTCTAGG
 196081 CTATGAGATA GAGGAATCCA AGGAAGAAGA TAAGCTACTT GGTTCCTCTA TAGGGTCTTG
 196141 TGTGTGCTCT CATGTGCTCT CTCTCTCTCT CTCTCTCTCA CACACACACA CACACACACA
 196201 CACACACACA CACACACATG AATACCAGAG CTATCACTTT CCCAGTCTAG TACTCATCTC
 196261 ATCCCAAGGG TTTTGTGTTG TAGTGGTTTG CTCATTTGTT TGTTTGTTT GTTTGCTTGG
 196321 ATTATTCTTT TTCTCTTTTT GCAGCTGAAG GGAGAATTTT CAGGCCAGCC CTTTGGCCAT
 196381 TAGAGTTACA GTGCCTCTAT TCAGGCTTCA TAGAGAGACC TGGGATTGAG TAGTGGGGGG
 196441 CTTTTATCCA GTTCAAAATA ATGCTATTCTC ACCAAGATGT ACTTTGAAAT AAAACAATAC
 196501 TAAACACAA AATTTTATTT ATGCTGAACA TTGAATCACT TTTTCTGTA TTTTGTGTAG
 196561 AAAGTTATAC ACACACAAAC ACATTTGCTC CTGCTTTGTT TATTGGCCCA GGGGTATGTT
 196621 TGGTAATACT TCATCAGGCA TGAGTAGTAC GTCTTGGAAG GTGTGGTCTA AAGCCTAGAC
 196681 TCCTATCTGC TTCCTTCAGC ATTCTCCAGT GTATCTGTCA TCTGTCTACC TTAGGATGGG
 196741 GTCTCCAGAA CTTCCATTCA CATTTAGAAG AGGGCAGCGG CTTTCTATGG AAAATATGAA
 196801 CTCTCATTCA TCTCTATTCC TTCTTCTAGC TATGGTCCAG CTCAGCTGTT TGGGAATAAAG
 196861 TATCTATATG AAGTCTGCGA ATGGTTCTCA GACTGGTTGA ACATTAGAAT CACCTGAGTA
 196921 CCTTCTAAAA TTCTTATTAC CCAGGGCATA TCTCAGAATG AGTACCACAG GGTAGGGATA
 196981 GGATTAGGGA TCATGATCTC TGGAGTCTGG TTTAGGCACT AGTGCTGTTT AAAACTACGT
 197041 TCATGAGGTG GAGGTTGCAG TGAGCCGAGA TGGCGCCACT GCACTCCAAC CTGGGCGACA
 197101 GAGTGAGAGT CTGTCTCAAC AACACAAAAC AAAAAAACC AACTACCCTT GTGATTTGAA
 197161 TGTCCATCCA AAATTGAGAA CCATTAGGTA AGGCCAAGCT GTATAATTA AGAGCAGTTT
 197221 TCATTTGTCT GGTGTGGTGG CAGCTTTTTG ATAAGGGAAG TATTGTTGCC ATCCACATAC
 197281 CTGAGCCTCA CTCCTGAGAA CACTGGTGTG TATGTTGCTA AAATCCCCA GGTGATTCTG
 197341 AGGTTCCITC CTGGATAAAA ACCACTGACC CTGGGAATGT ACCCACTGCC AATCTCCTGC
 197401 GTAAACCTTG GATACTGGGA AGCCTACAGT TGAAAATATT GGGCTTGAGA TCCTGAAACA
 197461 AATCTTGAT TTCAATTAAGA CTAATATTTG GTACAGTGCA GCAAATCAAG GGAATTTTGG
 197521 TGGCTGAGTT CTTTGTAGAAC TTTTGCATTG AAATAGGTTT AAGCAGCAAT AAGTTAAAC
 197581 TACAACCTCA GCTAAAGGAT TAAAGACAC GTGAGCTGGG TAGGATGAGG TCTAAGATTG

Figure 9 (Page 61 of 74)

SUBSTITUTE SHEET (RULE 26)

150/162

197641 GGTGTGGCGG CTCATACCTG TAATCCCAGC ACTTTGGGAG ACTGAGGTGG GTGGATCACT
197701 TGAGGTCAGG AGTTCAAAAC CAGCCTGGCC AACATGGTGA AAACCCATCT CTACTAAGAA
197761 TACAAAAAAA TTAGCTGGGC GAGGTGCCAG GCACCTGTAA TCCCAGCTAC TGGGGAGGCT
197821 GAGGGAGGAC AATCACTTGA ACTCAGGAGG CAGAGGTTGT AGTGAGCTGA GATCGCACCA
197881 CTGCACTCCA GCCTGGGTGA CAGAGCAAGA CTCCATTAA AAAAATAATA ATAATAATAA
197941 CAATAATAAT AATTCAGACA TATCCAGGCA TCAAACAGAT ACCTGGGGCA GATGAATAGT
198001 CTTGAGATTC AAGTCACACA TGAAATTTAG GTGGAATG ACATTGGAGA AATTTGAGAT
198061 TATGATGAAT GGAAATTTTT CAAAGAGGAA TTTCAGGCTC TGTTCTTGAG GGGATAGATG
198121 GACTTCCAAC AGCAATAACA CAGGATTAAT GAGGACTTGG GATGTTACAT AAATTAGAGA
198181 TGTTAGATGG ATAAAGAGAT AAAAGTACTC TCTCTAAGAA CATGGGACCA GAGATAGGCT
198241 CACTTCTAAC CATCAGATAT AACTAGCAGA CTAAACGGTC TAAAAATAAA AATCATGCCC
198301 CACTCCTGCT TAAGACATTT TAATTACTCT CAGTAACTCT TCAGTTTTTC TACTGTGTTA
198361 TCTTTAACTA CAGGGTTGGT CTGGGTGTGC AACACAAGAA AGCCTGGCAT ATACATGGAT
198421 TCAAGTGTAT GCCATGTACA GGTATCTTT CATGTACTAT TTCATGTATT CTTTTTCACA
198481 TCTGTTTTTT CCTTCATTGA AGTCAATGGC TGATATTAGA TTCTACTATT CATGTGTACT
198541 AGTTATATAT AATTGTTACA AAACAAATTA GCAAAACTT AGTGGCTTAA AGCAACACAC
198601 ATTTATTATT ACCTAAGGTC TGTGGATAGA AGTCTGACA TGGCTTAACT GGGTTCCCTG
198661 CTTCAAGCCT CATGTGGCTG CAATCCAGGT GTTGGCTGAG TCTGAATTCT CATCAGAGGC
198721 TTGATTGTGG AAATTTCCAC TTCCAAGCTC CCTCAGGTTT GTTGAAAAAT TCAGTTCTTT
198781 GCACCGGTAG AAGCTTCTTG GTAGAGGCTG ATTCAACTTC TAGAGGCTGT CTGCAGTTCC
198841 TGTCACCCAG GGTGGAGTGC AGTGGAGCAA TCATAGCTCA CTGCAGCCTT GACCTCCCAG
198901 AATCAATCTG TTCTCCCACC TCAGCATCCT GAGTAGCTGG GACCACAAGT GTGTGCCATC
198961 ACACCTGCCT AAAAAACAAA CAAACGAAAA AAAACCCCCA GAGAACTTTG TAGAGACAG
199021 CTGGTCTGGA ACTCCTGCGC TCAAGCAATT CTCCTGCCTT AGCCTAAAAG TTCTGGGATT
199081 ATAGGTATAA GCCACCATAC CTGGCATATG GCAAGTCTTG AGCAGGACAA ATACAGATGA
199141 TTTATGTCTG TCTTCCATGG TATTCTAGGT TATTGTTGAG ATGGTCTCTT ATTGCTTGT
199201 TCCATCTATT GATTAGATAA AACGTGTTC CTTCTGTTAT TTTTCAACAG TAGCTTTTAT
199261 GTGTCTCTCT TTATCTTAAA ATTCTAACCA AAGAGCTGCT CTTTCTTGG TGTACTTTAC
199321 CTTTGGTTGA TCCTTCTTAA CCTCTCTTG CCCTCTGGGG CCTAAGATGA GGGCTGTTAT
199381 CAGATGTGAG TCTATGGGAA AGCAAGCAAG AGGTTCTTCA GCCTCCGTTT AGCCTTAAAT
199441 GTCTAGGTAG AAATCAGTCA TGGCCCTTCC AATGTGGTAC AGACCAGATC ACAGAGACAG
199501 GGGTCTCAGC CAAGGTCTTG TGGCCTAAGC CTTATAGAAA TAATGAGTGT TTACTTACTT
199561 GGAGAACTCC CTTGGAATAT CTTTTTTGT GAACCTGAGG CAACTTTTGG TGATTCTTG
199621 ATGTCTTGGG AATCTTGGTC TAGAGCCATT TCAACCTGAT TTCTTTTCAT GTCAGTGGCA
199681 TTTTGTGACC AGATAGTAAA TAAGTTCTAT GATGTTCACT CAGAGAAATA CAATGACTTA
199741 TGATGTGAAG CTTCTGTGGT TCAGCCCTTA CTTCTATCTT ATTCCCTCTT ATCTGCATCT
199801 GTCTCCTGCT TGGGAACAAA AGTCTGGCTT CATTCTATGA CCCCCACGTT GAGTTTCTTA
199861 GTAGCACTTA CTTTTCAATT AGGAGTGTC TCACTTCTAT CCATCAGACA TAACTAGCCG
199921 ACTAAACAGT CTAAATATAA AAATCATGTC CTACTCCTGC TGAAAACATT TTAATTACTC
199981 CCCATCATT AATTTTTTCT ACTGGGTTAT CTTTAACTTC AGAGTTGGTC TTGTGTGCAA
200041 CACAAGAAAA CCTGGCATAT ACATGGATT CAGTGTATGC CACGTGCATG TATTCCTTCA
200101 TGTACTATTT CATGTATTCT TTTTACATC TGTTTTTCC TCTAAAATTT ATTTCTTTT
200161 AAAAATGAAA ATTTTGCATT TGACTAAAT TGTCAAATTT AGTCAAATTT GTTTAAACC
200221 ATTTTTAAAA TGTTTCCCGA AGTTTGGAGT GAAGTTAGTA CTTCAGAAAA ACTGTTTTGT
200281 ATTTTTTCATG TGACCTCAGT GCACTGCTGT GCATTTCCAT TTCTGCGTCC ACACACATTT
200341 GTTTTGAGGA AATATAGGAA CGACAAGATA AAGTTCAAGC TCCTGGACAT TGCATAAAAG
200401 ACCGTCATGA CCTGGTCTG TTAGCTTCCC TAGATTTCCC GCTATTTCCCT AAGTTGAGAT
200461 TTTTGGTTTG GATGCTTTGT GTTTTCTTAA AATCAAAATA GGTTTTTGCC TTTTATGATT
200521 ATACAGTAAA TAAATGCTAT TTGTGTGAAA CTTTAAACAA TACAAAAAAA ACCTAAGGAA
200581 GAAAGTCAGA TTCATCTAAA AATCCTTTGT GCCAGAATTA ACTACCTTAG TTATTATTTT
200641 CTCTATCTCT CTCTCTCAAT GTATATTTGG TGATAGTATA GGGGTGTGTG TAGTGTGTGT
200701 GTATGTATAT ATCTGTTTCT ATTCTGTAT GTGGATGTGC ACAACGCATC CTGCTTTGTA
200761 CACTACAGTA CTAGCATTTT TCTAATGTAA TTCAATATTG TTGAAAACAT TTTAAAAAG
200821 CTTGTATATA TACACACACA TACACATACA TGCATGTATG TACATATACA CATAACAGACA

Figure 9 (Page 62 of 74)

SUBSTITUTE SHEET (RULE 26)

151/162

200881 AAAATGTATC CTATGTATAT TCACACATGT ATACACACTC ACACGTACAT AGAGTTTTTAC
200941 ATCCATAGTT TATAAATGTT GCTTTTTTTT GGTACCTTT TTGCTAAGTC TTACACTTTT
201001 TTTTTTTTTT TTGAGACGGA GTTTTGTGT CATTGCCAG GCTTAGTGCA GTAGCGCGAT
201061 CTCACCTCAC TGCAACCTCG ACCTCCCGGG TTCAAGCGGT TCTCCTGCCT TAGCCTCCTG
201121 AGTAGCTGGT ACTACAGGTG TGCGCCACCA TGCCTGGCTA ATTTTGTAG TTTTTTATA
201181 GAGACGAGGT TTCACCATGT TGGCCAAGCT GGTCTGGAAC TCCTGACCTC AAGTGATCTG
201241 CCTGCCTCAG ATTCCCAAAG TGCTGGGATT ACAGATGTGA GCCACTGCAC CCGGCCAAGT
201301 CTTACACATC TTTTTTTTAC CACTAACTG TTTACCCAAA CCTGATAACC CAAGTCAACA
201361 GCTATTATGG CTCACACAAT CTTATGTAAA CAAAGATACA GATATATAGA ATTTCTTGA
201421 TTAATATTCA GAAAAAATG GAGTCCCTTT ATACGTCCTT AGTATCTGCT TTACTCATT
201481 AAAAATGTAT TACATTATAT GAAAGTATTC AGGTCAAATG TTATAGATGT GATTCAATTCT
201541 TTTAACTGT GTTATTTTTT TGCAATGACT ATGTATCACA AAGTACTCAG TCTTCCACTG
201601 ATGAAAATTT GGGCTATTTT CAGTTTGTCT TCCATTTTTT TTTCTCCTC TTGGATTTT
201661 ACTCAATGTG TTTACTAATT TAGGAAGAAT CAATAGTTTT TATGGTATTA CTTCTCCCAT
201721 TCAAGAATAT AGCATATGGT ATAGTATAGT AGAGTACTTA GTTTAATTTA GCCAGATCCT
201781 GTTTTCTGCC CTTTAATAAA ATTTCTATCAT TTTCTGCCTT TGAGTCACAT TTTCTTGT
201841 CATATAATTC TTAATAAATG TATAGTTTTT ATTCTAAGGG AACATAAAAA CTTCTTTCCA
201901 TTTCTATTCC TGTCTAGTTA ATTTCTACTAT TGGGAAAAGT AACTGTTAAA AAAAATTTCT
201961 ATCTTTCCAG TCAGTTCACC ACATTTCTCT TATACCTTTG TACTTTAATC CCCAGTCATG
202021 TTGAACACTT CTTATTCCTC ACACCAAGCC TCAACGGGTT TGCTCTTTCT GGAAGGTGCT
202081 TCCCCTGTAT TACTGACTTA TTCATACCAC ACATGGAGAC TGGCGCAGCC CTGTTCTGCC
202141 TGGGAAGCCT TCCCCTGATA CCCCTAGTTG GCAGGAGTCT TCATTGTTC TTTCTAGTC
202201 ACCTGTGCAA GTTTGTATTG TTCATGTTTA TCATCCTTCA TTCTAGTTGT CTGCTCTAT
202261 GTGTGGTCTC ATTCAGTGA CTCTGAAGTC TTATGAAGTC ATGTCATGGG TCAGATCTTA
202321 ATAAATTAAT ATTGTGCGAA GCTAATGTCA TGTCTAGAAT ACAGAAAATT TATCAAAAA
202381 AAATATAGTA TGTTGGCTGG GCGCAGTGA TCAAGCCCGT AATCCAGCA CTTTGGGAGG
202441 CCGAGGCAGG AGGATCACAT GAGGTCAGAA ATTCAAGACC AGCCTGGCCA AAATGGTGAA
202501 ACCTCATCTC TACTAAAAAT AAAAAAGTA GCCAGGCGTG GTGGTGCCCA CCTGTAATCC
202561 CAGCTACTCA GGAGGCTGAA GCGGGAGGAT CACTTGAACC TGGGAGGCAG AGATTGCAAT
202621 GAGCTGAGAT CATGCCACTG CACTCCAGCC TGGGCGACAG TGAGACTCCA ACTCAAAATA
202681 ATAGTAATAA TAATAATAAT AATTGTATGG AATTGAAGTC CTCTGATTGG AAATAGCTGT
202741 TTTTTAAAA ATTATTATTT TTAAAGTTCC TGGGTACATG TACAGGATGT GCAGGTTTGT
202801 TACATAGGTA AACGTGTGCC ATGGTGATTT GCTGCACCTA TCAACCCATC ACCTAGGTAT
202861 TAAGTACAGC ATGCATTAGC TCTTTTACCT AATGTTCTCC CACACCCCCA CCCCATCCTC
202921 CCCCACAGG CCCAGTGAG TGTTGTTCCC CTCCCTGTGT CCACGTGTTT TCATTGTTCA
202981 GCTCCCACTC ATAAGTGAGA ACATGAGGTG TTTGGTTTTT TGTTCTGCTC TTAGCTGTTA
203041 ATGTCAGGCC AGAGAGGCTT AAATTTTTTA GGATCTCTGG ACTTTTCTTC TACATTACTC
203101 TTGATGTTTA TAAATGTTAC AACTTCTTTA ATTTCAATTA ATGTATACCT TATTGAGTTG
203161 ATTTAACTGA GTTAACTTTG TTATATGAAA ATCATGATTG GGAGTGAGGG GGTAAACCA
203221 GCTACAGAGA TCTTGATTGT TGGTGGTGAA GCAATGCAAG AATTCAATTC TTCAGTAAAC
203281 TAATGTTTAT TAAGCGTGTA CTGTCTTAGT CTGTTCCAGC TGCTGTAACA AAATATCATA
203341 AACTGGGTGA CTTATAAACA ACAAATAATT TATTTCTTAC AGTTCTGGAG GTGGGAAGTC
203401 TAAGATTAAG GCCCTGGCAA ATTTAGTGTG TGGTGAGGAC AGGTAGCCAT CTTTTGCTG
203461 AGTCCTAACA TGGCAGAAGG GTTGAATAAA CTTCTTGGG TTTCTTTTAT AAGGACACTA
203521 ATCCTAGTGA TGAGGTTTCT GCCCTCATGG TATAACTACT GCCCAAAGAC CCTCCTTCT
203581 AATATTATCA CTTTGTGGGT TAGGATTTCA ACATGAGTTT TGAGAGGATA CAGACATTTG
203641 GATCATAGCA CACACCATAG GACAGACACT GTGCCAAGAA TTGTGGATAT AGTGATTCTC
203701 AAAATGAACA AGATCCCCCTC AGAGAGCTTG CAAAATCCAG CTATAAAATT ATGCTTTTTA
203761 AACAAATTAT GCAGTTTGAA AAATCTACTC TGAATCTTAC TTGTGGCATT GAATACTTTC
203821 GGCCACTCTT TCCTTATTAT ATTAATATT TACTCTTGT TGGGGGATCC AGTCTCACCT
203881 ACTTTTTCTA CCAGAAGTGG TATCAGCTCA TGCTCTGCCT TATGCAAAAT AAGAAAATAT
203941 CATACCTTTT GGGTAAATTA AGCCAAGAAA GTTCTCCTTT CTTCTCTTTC TCTCTTCTT
204001 TCTTCTCTC TTTCTCTTC TTTCTTCTC TCTCTTCTT TCTTCTTTC TTTCTTCTT
204061 TCTTCTTTC TTTCTTCTT TCTTCTTTC TTTTCTTTC TTTCTTCTT TCTTCTTTC

Figure 9 (Page 63 of 74)

SUBSTITUTE SHEET (RULE 26)

152/162

204121	TTTTTCTTTC	TGACAGGGTC	TTGCTCTATT	GCCTAGGCTG	GAGTGCAGTG	GTGCAATCTC
204181	AGCTCACTGC	AGCCTTGAAC	TCCAGGGCTC	AAGCAATCCT	CCTGAGTAGC	TGGGACTATA
204241	GGCATGTGCC	ACAACATCAA	GCTAATTTTT	GCATTTTTTT	GTGGAGACGG	GATCTCCCTA
204301	TGTTGCTAAG	GCTGGTCTTG	GATTCTGGG	CTTATGCGAT	TCTCCTGCCT	CAGCCTCCCA
204361	AAGTCCTGGG	ATTACAGGCA	TGAGCCACTG	CCCCTGGCCA	TTATAACTAT	TTTCATTGGC
204421	TTATCAGGCA	CATGATAACT	ATAATAATC	AATAACCAGA	ATTTTTAAAT	AAAGAAAGGA
204481	AGGAATTGTT	TCAACTCTTC	CTGCTACCCC	TCTATCCCTC	AAAAGGGTAG	GCTGAATGTT
204541	GTCCTCCAAA	GATATCCATG	TCCTAATCCC	CAGAACCTGT	AAATATATTA	CCTTATATGA
204601	CAAAAGGGAC	TTTACATGTT	TAATAAGTTA	AGAATTTTGA	GATGGGCAGA	TTTTCTTGAA
204661	TTTTGCAGAT	GGGCCCTAGT	GTAATCACAA	GGGTCTTAT	AAGAGACAGG	CAGAAGAGTC
204721	AGAATAAGAG	AAAAATACTT	CAAGATGTTA	CACTGCTGGC	TTTAAGGTGG	AGGAAAGGCC
204781	AAGAGCCAAA	AAATGCAGTG	GTCACTACAA	GCTGAAAAGA	AAAAGAAATG	GATTTTCCCC
204841	TAAAGCCTCT	GGAGGGGGCA	CAACCTTGCC	AATACCTTGA	TTTTGGCTCA	GTGAAACCCA
204901	TTTTGGACTT	CTGACCTTTA	GAAGTGTAAA	TAAATAAATA	ATTTTGTGTT	GTTTCAAGCC
204961	ATCACAGTTG	TGGTAATTTA	CTACAACAGC	AATAAAATAG	AATTAAATAC	AGAGATCTGA
205021	GGAGTTGAGT	AGGATAAGCC	TACTCCAGCA	GGTTATTTTCG	GGAGTATGGT	GAGACTCACT
205081	AGGATGGCGG	AACTCAATTA	AGGAAGTCTG	AAGCTGATAA	GCCAGAGAGG	GAAGGCTCTC
205141	ACTTCATTTT	ATAAGGGTTG	CGTCACACTA	GGAAGATCCA	ATAGCAACCA	CAGTCTCAAA
205201	ATTAATGATT	ACAAATAGGA	CACAATTCCA	AGAGTCGGGA	GCCAAGCAGA	AAATGGATTA
205261	GGGAAGACAT	GGATGATATG	AAACAGGAAG	GAGGGGTACA	AGGCAGCTTC	CTGGGAAGTT
205321	GCCAGGGCAG	TCACAGTTCA	CATTCAATAG	GCTGTGGGCA	CCAAATGCAT	ATGGAATACT
205381	TAGCTGACTT	AACTGAAGTC	CTGAAGAGGA	ATGAACACCT	CATTATTTGA	GGAGCTACTA
205441	CCAATTAGAA	TATGTATTTT	ATTTGTTCAA	TAACCCCATG	AGTACAGTAA	CACAATCCTT
205501	GCTTTACTAA	AGCGGAAGCC	AATTCAAAGA	GGTTCAGTGA	CTTGTCCAAG	CTCAGGGAAA
205561	ACACTAGGAA	GTGAATATGG	GTCTGACTCC	ATCACTGATT	TCAGGAGCCC	TGCCCTTTCC
205621	TCCACACCAT	GCCCCCTTGC	TTTCAGAAAA	AAAGGCTTGT	TGACTGAATG	GTTGTATGCA
205681	CAGTTCAAAG	CAGAAACACA	CGATGACATC	TTTTGAGATA	CTCTAACAGT	GAGAACTTGA
205741	AAATGAAGTT	AAAAATTAAG	CGGCAAAACC	AAGCCGAGGC	TTTCTGAGAA	AGTGGGGCCA
205801	AACCTGTTGC	CGTCTGACTG	CCACGTGGCT	CACTATTTAT	CCCTGTAAAA	ATCTGCAAAA
205861	GTATTTGAAA	GGGAAGAAGG	GACAGAAAAA	TCCCTCCTTT	TCCAAGTTAG	CCTTATAGTC
205921	TAGGGCTTAA	AATACTGGTT	TAATGGTGAA	GGTAAGTGCT	TTTCTTCTTT	TTGGGTAGAA
205981	GGATTATTAC	TAACCTACCA	AAGGTCCATT	AAGGGGAGGG	AACAGTTTTA	GGAGAAGTCA
206041	GAGAAAAGAC	ATTAACAGCA	ACATAAGGAT	CTCCATCTGG	TAATATTGCC	TAATTCCAAA
206101	ATGAAGAGAC	TCTCTGAAAA	AGATAACTGA	TTCAATGAAG	ACCCTAGGGC	AAGGCTTGAG
206161	AAGCCACTGG	TACCAATGGA	CACTGTGGAC	AATGGTCATT	TCTCCAAGGA	CGCTGTGAGT
206221	ATTAAGTGTG	ATGCTGTGAT	TAGTCAGACT	GGGATTGGCT	GTGGAATGAA	ATACTGATCA
206281	GAAGTACAAA	GATTTGTGTT	TGGGACTGTG	GCTAACGAGT	CTTTTCAGAC	TTCTATATGA
206341	ATTTGAAATG	GTCTCTCAGG	AAAAGGAGAA	CATGGCCGGG	CCTGGTGGCT	CACGCCTGTA
206401	ATCCAGCAC	TTTGGCAGGC	TGAGCGGGC	AGATCACTTG	AGGTCAAGG	TTTGAGACCA
206461	GCCTGGCCAA	CATGGTGAAA	CCCTGTCTCC	ACTAAAAATA	CAAAAATTAG	CAGGGCGTAG
206521	CGGCGCGTGC	ACCTATGCGC	ATGCATAGTG	CGCGTGCCAG	CTATTACAGAA	GGCTGAGGCA
206581	GGAGAATTGC	TTGAACCCAG	GATGTAGAGG	TTGCAGTAGT	TGAGATCATA	CCACTGCACT
206641	CCAGCCTAGG	TGACAGAGTA	AGACTCTGTC	TCAAAAAAAT	AATAATAATA	AAAGAAAAGG
206701	AGAACATGAC	CAAAGTTATG	AATAAGACTG	AAGGCAAGAA	AATTGTACGC	TTGTAGAGAT
206761	CACCTAGCTT	GTTGCCCTCA	TTGTACAGCT	AAGAAAAGGC	ACCCAGGGAC	ATTGTGGTCA
206821	GCACCAATTT	CTCAGAAAGA	TAGGCAGATG	ATGAGAGGGC	CCTCAGTTT	TCTAACACTG
206881	AAGGAATTGC	TTCTATGTTT	TCTGGTGAAC	TCCTCCCCAC	TCATCTTGAG	GATTCCAGGC
206941	CAGAAGAATC	CACTTTAAAA	AAGAAAACATT	TAAAAACCAAT	TTAACAACCA	ATCAAAGGCA
207001	CTTTTATAGA	AATACATTTT	ATTTGCTGTT	GGCCTGTATT	TATGGATCTG	AGAGGGCTAG
207061	ACTGCCAATA	TTGTGACTGT	TTATTATTAT	TGCTGTTGCT	AGTATCTAGA	ATATTATACA
207121	ACATATAACA	CTTTGCAATT	TACGAGGCAT	GTCTCATACT	TTTGTTTTCA	CTCCAACTG
207181	CCCAGTGAAG	TAACATTATC	CCAATCTTTC	CTATGAAACA	GTGAAAGCCC	TAAGAGTTTT
207241	TGAAACTTTA	CCTGGTTTAC	TCAATTTGGG	AATGGCAGAG	CAGAATTCAG	TCCTTGAATA
207301	TCCTCCCACT	GCAGGTTTAT	GCTCTTTGAT	CTAGGTGTAA	CATTTACTCT	GAGTAACTA

Figure 9 (Page 64 of 74)

153/162

```

207361  GGACTCTGGG  CTAACAGAGA  TGAAGCAAGA  CAGGCTGGAT  ATTAGGAGAA  TCTAAGAGCA
207421  ATCTAACGAC  CATTATAATA  AAATCATGAG  TTCTAGACTT  AAAAAAGGG  AAAAACCTGT
207481  TTTTTGCTT  ATGCGTATAC  CATAATATTT  ACATTATTTA  TTTTTTCTC  AAATCCAACC
207541  TATACGGTGT  CAAGTAATTT  TTTTAATAT  AACATTTTCC  TTTAACTTAA  TTTCAATTCA
207601  TTTTCTGTG  TCTACTTACA  ACTTTGGCAC  TAGAATTCAC  AATTTTTTTT  TAGAGGTATA
207661  TCTCCTTAAA  GGAAGGGTT  CTGACACTGT  TACATGTTCT  CAATTGTTTG  CAAATAGGTT
207721  AATAATTATT  CCAGTGCTC  TAAGTACATA  TCAACCATGC  CAGTGTTTCA  CCTCCATAAT
207781  TTTATTAGCT  TCTGTGCTTA  TTTTGGAAAA  ACATTTCCCA  TTACCATGAA  AGACCTCAGT
207841  TTAGGATGGT  TTGGTATGTT  AGCCTGATTT  CTGCATTTCG  CTCATGCAAA  GGAAAAATAGG
207901  AAACGAAGAA  CTGAAATTAC  CTATTGATAC  AAAATCAAAG  TAGCATTTGA  AACCATAAAA
207961  CTTAAGTAGG  GCTTTTCATC  CTTTCTCGTT  AGACAGCAAC  AGAGAATGGG  AAGAAAAACT
208021  AAAGTGATGG  GTTGTGATA  CAATTCAGT  AACATAAAGA  GCAAGGAGAA  GTAGTTTGT
208081  TGTGTTTATG  TTTAATATTC  AAAGCTCAAC  CTAAGATAT  TTTTCATTAT  CAAACTTCCT
208141  TCTAGAATAA  ATGATTAAAA  CTTGATTTAA  AATATACAAA  TTCTCCTTTA  TAATACCTCA
208201  AAATGGAGCT  ACCCCATTGA  GTTTAAGCT  TGTGATTAAA  ATATTACGAA  AACAAAGGGG
208261  AAGTTGTAAT  AGGTAGAACA  AGCAGTAGTC  TAGGCATTAG  GGGATCTGGT  GCTGGCTCTG
208321  TGCATCATGT  GGTTCAGGC  AACTTTTCAA  ATTTTCTACG  CAAATTTTCT  TATCAATAAA
208381  ATAAACAGTT  GGGCCAGAGG  ATCTCTGAGT  CTCTTTCAGC  TTTTCAGTGT  TATAAGATTG
208441  GAGAAGTTGG  TGGGAAAGCT  TTAAGTGGAG  TGTAAGTAAT  TGCAGCTGCA  TGTACAGTTA
208501  AAGAGTTGCC  TTCAGCCAAG  CCACGGGATC  TTGCATAAAA  AGTGAAATCA  AATAGAAAAT
208561  GGTCCAAACT  CTGGGTTTGA  CCACAGTAGA  CTTCAGCTAG  GATCTGAGTG  TAGAGCAATG
208621  AGCTGAACTC  CTGATATCCA  GATGTTAGCA  AGACTTGGAG  GCCTTCTAAG  GCAGAGCAAC
208681  AACCAGTATC  TGTCTGGTG  CTGACCTGAT  CTTACTAGCA  ATTGGGCCTC  CATTGGGGTC
208741  CATTGTACAA  AACAACAACA  ACAACAACA  TAAAATCTCC  AAACACCCAA  AATTCAAAAT
208801  TTAGATGGAG  AGATACTATT  CCCAGAATTC  TAGAGATATT  TGGAAAGCAG  AAACTATAC
208861  TTGCCATGCT  GATGAAGTCC  AATTATTGCT  CTTTTAAATA  CATTTAGCTA  CTCTGAATA
208921  TAAAATGAGT  ATCTACTAAT  TATTTACAAA  ATCACTTGGT  AAATATAGAA  AGTCACAAAG
208981  AATGAAGTGA  TCATCCTGTT  TTGTAACCCA  GAAATAGTCA  TTACTGGCAC  TTGTGTGAAT
209041  CAGTTTCTAT  TCCTGTATGT  GGATGTGCAC  AGCGTATCCT  GCTTTGTACA  CTAGAGTACT
209101  AGCATTTTTC  TAATGTAATT  CAATATTGTC  GAAAACATTT  TAAAATAGCT  TCCATCACAA
209161  TAATCTATCA  AATTGACTTG  CCAGACTCTC  ATTATTAGGT  TAATTTATCT  CTAACATTAT
209221  GCAGTCATGA  GTAATACTAC  AAAGGATATT  TTTGGACACA  ATTTTTCATC  TATGCCTTTC
209281  TTTATAATCC  TTCATCCTAA  GGTACAGAT  TATGAATATC  TTTAAAGTAC  GGACAAGTCT
209341  TTTAAATTTT  GTGTGCAAAA  ACAGTGCAAA  GCCTTGAATG  ATAAATAGA  GGTTTGATAT
209401  ATGTGTTTTT  TTGTTTGTGT  GTTTTGAGAC  GGATTCCTGC  TCTGTCCCCC  AAGCTGTAGT
209461  GCAGTGGCAC  GATCTTGGCT  CACTGCAACC  TTTGCCTCTT  GGGTTCAAGC  AATTATCCTG
209521  CCTCAGCCTC  CTTAGTAGCA  GGGTCTACAG  GCATGTGCCA  CCACACCCGG  CTGTTTTTGT
209581  ATTTTITAGT  GAGATGGGGT  TTCACCATGT  TGGCCAGGAT  GATCTCGAAC  ACCTGACCTC
209641  AAGTGATCCA  CCCACCTCAG  TATCCCAAAG  TGCTGGGATT  ACAGGTGTGA  GCCACTGCAC
209701  CCGGCCGATA  CATGTGTTTT  TAAAGTCACA  GAAATTTTCA  ATGTCTTGAA  GGATTTTAAG
209761  CAATTTAAAA  AATAAAGTCA  TAGAAGCTTC  AATTTAGGAA  TGAATGGAAA  ATTGATGATA
209821  TTCTTAGGAT  ATGGATTTTT  CCTAAAAGAA  ACAAATGTAT  GCATCCCCAA  AGATAATTTG
209881  ATTAGTATAC  AAATATTAAA  TTAACATGT  CCATATTAG  AGCCATGAAT  TCTCTTTGCC
209941  TGTACAATA  GCTGGATTTA  TTCACAATTG  TAGTAATTAG  TCCCTGTTCA  TTATAATTTT
210001  CTAGGTGATA  TGAAGACTTT  GTCAGTCCAA  GCAAGTGTC  ACATTGTGTG  TAGCAAACAT
210061  GAGAATAAAC  ATTTTAAACT  TTTAAATGTA  ATACATATTA  GTGTTATGTA  ATGTCATCCT
210121  TCATGTTTGA  AGGCACATGG  AACATTGTTC  TGGTGGTACA  GAGGGGAGAG  AAACACCATC
210181  AGAATGAAAG  GAAAGACCGC  TCTGGAACCT  TCCTCCTTAG  CTCTTGAGCT  TAGTTTAATT
210241  GTCCTGTCTT  ATGGTCTGCT  ACAAGCAATA  CCACTCTTCA  CCTTCGCATG  CTTCTCTGTG
210301  GTTTGATAAA  GTACATGCAA  TTTTTCATTT  AATTCCTCCA  GCTGCACTAA  GAAAGGAGCC
210361  TTATCTTTAT  TGAACAGATG  AGGAAATGAA  TGATTAGAGA  ATTTAAATGA  CTAGCTCTAG
210421  GTCACACAGC  TGGAACTTAC  AGCCAGATTT  CCTTTAACA  ATCCTGTAAC  CAAAAGCATA
210481  CCAAGTAGTC  CCCATAAAAT  GTAAGTTATA  GAGCTGTGTT  GGGTCAAAAC  TTTTACTGAT
210541  GCTAAGAGGA  GGCAACATTA  ACAAGGGGAA  ATTATTTGTG  TATTATGTTT  TGGATTATGT

```

Figure 9 (Page 65 of 74)

SUBSTITUTE SHEET (RULE 26)

154/162

```

210601 TCTCTCCATA GATAAAAGAC TGTCGTAGTA AAAGAGATTC AGGGCACAGG GAAACTCCAC
210661 CACAAAGCGT GGTACCATT CCCACAGAAG CTAAATGGAC GGAAGCCTG CCACCAGGAA
210721 AGGTAAAGCC ACTGCTCTTG TTTGCAGGCT ATGTTAATAA GCTGAAGCTT ATTCCGACAC
210781 ATTTACACAT CTCTGCATCA CACTGACCCT TCGTAAAGAT ACTCCCAGTG TAACATTGGA
210841 GCCAGTCCA GCCCCTGATC CTGTTGCTTT TTCCTTAGCC CCATGAAATC ATCTGCGAGA
210901 AATTAAAGCCA AATAAGCAAT AAATCCTGGG ATCTAGGGAG TGAATAAGT TTTGGGAAAG
210961 TCTTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
211021 GCGATCTCGG CTCACTGCAA CCTCTGCC TC CGGGTTCAA GTGATTCTCC TGCCTCAGCC
211081 TCCCAGTAG CTGGACTAC AGGCACACAC CACCATGCCC AGCTGAATTT TTGTATTTTT
211141 AGTAGAGATG GAGTTTCGCC GTGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCGTGATC
211201 CACCGGCTC GGCCTCCCAA AGTGCTGGGA TTACAGGCAT GGGCCACCAC GCCTGGCCCG
211261 GGAAAGTCAT TTTAAACCAA CCTATGTATG AATCCCTACT ATAATATTCT CACCAAGCGG
211321 CTGGCTCTTT CTCCTGAGCT TGGAAACCTC CAGTAAATG GAAATAATTA TTTCCAGAC
211381 CACCACTCTT ATCTGTGAGC TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
211441 ATCTGTGTCT TCACAGGTTT TCTCTTTCTT TCACCTTAGT GCTTTTCTTC AAATAAGCAG
211501 GAAAAATCCA ATCTATCATG CACATGGGAA CCCTTCAAT ATTGGTCTGT GGTGTTTCCA
211561 TTTTATGGGG ATGCTTTTAA AGAAAAAATT TGTCTTTCA ATATATTGAA TATCTTCCAG
211621 CACCACATCA CTGCAAGCT TTGTAAAAAT AGTTCTACAT ATTAATTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
211681 AGATTGAGTC TCATTCTGTC ACCCAGGCTG GAGTACAGTG ACATGATCTT GGCTCATTTG
211741 AACCTCTGCC TCCTGGGTTT AAGTGATTCT CCTGACTCAG CCTCCCGAGT AGCTGGGATT
211801 ACAGGCATGC ATCACCATGC CTGGGTAATT TTTGTATTT TAGTAGAGAT GGGGTTTCCAC
211861 CATGTTGACC AGGCTGGTCT CAAACTCCTG ACCTCAAGTG ATCCACCTGC CTTAGCCTCC
211921 CAAATGCTG GGAATACAGG CGTGAGCCAC TGCACCCAC GTAGTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
211981 AGTTGAACAT ATGTGAAGGC AGGACCTAGT GACACATAGC AATAACATTT CCAAGTAGAC
212041 ATTACACTAG GGAATTAGTC AAAGTGCTCA TTTAAAGTAC CATCTCTCAA ATGTATTAAA
212101 AGAGAATCCT TGGATGTGCA ATACCTTAAT TCAAAGGCAG CTCGTTATGT ATAACTCTC
212161 AAGCTTTGTG ATAAACAAAT GTGCATAACA GATGGGACTA TTGACTTACA GCCCAGGGAA
212221 TTTTATTGAC GCTGAGAAGG TTATGTGACT GGCTCTGCCA CTGTCATCCC CATTCACTTC
212281 ATTTTGGAGC AATATGACAT AAATGCCTTA CATGTGGGTT TTCTCTATTT ATCATGTGTT
212341 TCCTATCCCC TTGAAAGATG GCCATATTTG CTTTACTTGG TTATAAGATC CCATATTCGC
212401 TGCTTTGAAG CCAACCAAAT AATTGACAA AGTGGGTTT TAGTGCTGGC TATTTTGGTG
212461 AAAAAAGAC AATGAGACTT CATGTGTCAT CCAAAGTTCT ATCAGATCGA GCTGTGAGAG
212521 AAAGGAAAAG AAAGGGGTCT CAGTCAGGAT GCTCACTGCA TACATCTGTG TTGTTGTCTA
212581 GGTCCAGATT TCTGTTCAAT ACGTATGGG CTGGCTCTTA TCATGCACTT CTCAACTTC
212641 ACCATGATAA CGCAGCGTGT GAGTCTGAGC ATTGCGATCA TCGCCATGGT GAACACCATT
212701 CAGCAGCAAG GTCTATCTAA TGCCTCCACT GAGGGGCTG TTGCAGATGC CTTCAATAAC
212761 TCCAGCATAT CCATCAAGGA ATTTGATACA AAGGTAAGTA TGATGGAAAA TAGGGCTCTT
212821 TGTTGAGAGA AAAA ACTTTG AAAGGAAGGC ATAGATCTTG ATTCTGTGGA GTATGGAAGT
212881 ATACATTTCC AATGACAAAT TAAACTGAC TGGAACTATT TTTCTTTGAG ACATTGCTTA
212941 CTTCAATAAT AAAAATAAGA TTTCAATTGAG GTTATTATGA TTATAAGGTG GGGGAAGTGT
213001 AGAGTTAAAT GTGAAAAATT TAAAAATGGA ACAGTTTATG TGATGTCTTC AATGAAAAAC
213061 TAGGTATTAC CTGGGCACAT TCTTATAGGT TACTCAATCC TATTCAGTTC TCTGCCTGTT
213121 TTATTGTTTC TGAGCAATT TATATCCCTG TAAATTCTAT ATAACCAATA GAAATGCAAA
213181 CGATTCTTGT CCATAGCTTT GCAAATAAAT TTTGCCAAGA GAAAAATCAG TTAATACTTT
213241 TCTCCACTCA CCTCCAGTT GAATTAGCCA ATTTTGCTGT TTGTTTGT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
213301 TGAGATAGAG TCTTCTCTG TCATTGAGC TGGAGTGCAG TGGCATGATC TCAGCTCACT
213361 GCAGCCTCCG CCTCCCGGT TCAAGAGATT TTCCTGTCTC AGCCTCCCAA GTAGCTGGGA
213421 GTAAGGGGGC ATGCCACCGC GGCTGGCTAA TTTTGTATT TTTAGTAGAG ACAGGGTTTC
213481 ACTAGGCTGG TCTCGAATC CTGACCTCAG GTGATCCACC CGCCTCGGCC TCCCAAAGTG
213541 TTGGGATTAC AGGTGTGAGC CACTGTGCCA GGCTCTGCTG TATATTTAAA GTCTATTTCA
213601 GCATTGCTTC CTGCTTGTGT TATGCGTGAT TCTTTGAGTT TTCCTTTGAA CCAGTTATAA
213661 CATCTTACTT ACTTCCTCCA TTAATCAATG AGTTAAATAA AATCTTTGTG GTATGTTTAT
213721 TTTACATTTA TATGAAAACC ATGAATTTAC CCAATTAAAA AAATTATCCT TTAATTATC
213781 TTGTACTGTA CATTTCCCAT GTCATCCCTA TAATTCATGA TTAATGATT TATTACATTG

```

Figure 9 (Page 66 of 74)

SUBSTITUTE SHEET (RULE 26)

155/162

213841 GACCTAGCTT ATTTACAATG AGTACATAAA TTTATTGTCT CCAGTCTTTC CTCCATTATC
213901 CCGTCTACAT ATCCCACTG AGTAGATTCA CTACTCAGGA ATCTTGGACA CCTTCAAGTT
213961 GCCAAACATG CAGTGTTAC TGGACATGCT GTGTTCTTTC AGAATTGGGG CCTGCTTCTC
214021 AGCACACTCA CATCTGCTAT CAATGACCCA TGGAAAAGTTT TTGCCCTGAG CAAGCCAGAG
214081 TCCCTGTTAG TTTCTTCCAA ATGCTACAAG TTCACCTTTG CTATTTTTTC CGATGAGATA
214141 AAATTTTCCT TTTTGACTTT CTACAAATCA TAGTCATTTT TCAAGGGATA GTTCAAGTAT
214201 TGCTTCCTTT CTGGGACCTT CCCAAATTAT TATTTTCTCC TCTCAAAGTC TCTGTTTTAT
214261 TTATGTTTCAT CCTCAAATCT TGATTCTCAC ATGAATCATA TACCTTGTAT TATTTATAGT
214321 TTTTTTGAGT AGGTAAAATA TTTTCATATT TATATTCTTT GGCTCTCTAC TTTATAGCAT
214381 GATGCCAGAT ATTTAGGGGC TTACTGTCAT TTATTTTTTA TTTTATTTTA AAATCTATTT
214441 TATTTTTTAT TTATTTATTT TAAAATCTAT TTATTTTTAG GTAAATATTC AGGTAAATATA
214501 ATTTATGTAA TTATTTAGGA ATTTTAGGTA GTTATTTTAA AATAATTCAA ATTATTTATT
214561 GAGTTATATC AGAAGAATGT GATCTTATTC ATTTGTAATA TGTGTTTTAG GAACTCAGTT
214621 CAGCCAGGGC AGACCATAAT TCCCAAACCT GACTTTTCTT TTTAATTAGG CACTGATTTT
214681 GGTAAAGAGT TCAGTAAAGT TTTGTGTGTG TGTTTTAAAA AATTCTTTGA TATAAGAGTC
214741 AAGATGTTAC TCAACTTTTA CTAGAAGCAA AATAGAGGAA GTGCTTTCAC AGATGAAATA
214801 TCTCTCAATG TTTTCTTCCA TTTACTTCTT CCTATTATTC ATCTATATAA TCATTTTCTT
214861 TACCTCTTTT CTTCATTTCT TCTGTTTTTC TCTCCTACTA AGACAAGCAA ATTAGGGGTA
214921 TAATTGGTTA TTTGGGAAGG TAGGAAGAAT ACAGAGAGAA ACAAATATCA ATATTTTATA
214981 CTAGGGTCTC ACTAACCTCA AGCAACTCTG ACTGTAAAGT AGATTTTCAT AATAGGACTT
215041 CTTGACAAAG AGTTTTCTTA TTTTCCCCC AGGCTCTGT GTATCAATGG AGCCCAGAAA
215101 CTCAGGGTAT CATCTTTAGC TCCATCAACT ATGGGATAAT ACTGACTCTG ATCCCAAGTG
215161 GATATTTAGC AGGGATATTT GGAGCAAAAA AATGCTTGG TGCTGGTTTG CTGATCTCTT
215221 CCCTTCTCAC CCTCTTTACA CCACTGGCTG CTGACTTCGG AGTGATTTTG GTCATCATGG
215281 TTCGGACAGT CCAGGGCATG GCCCAGGTAT CCAGATACTT TCTCATTCTT GGTGGGATCC
215341 AGATTTCTGA ATTCTACAAA ATATCAAAGG TCTTAATGAT TTTCAATTCA GGAATGGCA
215401 TGGACAGGTC AGTTTACTAT TTGGGCAAAG TGGGCTCCTC CACTTGAACG AAGCAAGCTC
215461 ACCACCATTG CAGGATCAGG TAAGTGTGCA CAGATGGGTC ATAGCTTTGT CATCTGTTCC
215521 ATCCCACTGT GCTTATCTT CTATGAATCA AATGGTTTGG GGAAGAGAGA GAAAAAGTAC
215581 TGCTGAAAAA TTCAACAATA TAAGACACTT GCATCACAAA TAGGAAAGAT GCATCTGTGC
215641 AGTAAAGACA TTGAAGCTTA GAAGTAGAAA AAACCATTGT GAGCTAGGTT TCAGCTCAGA
215701 AAAGCCTTAG TAGTCAGAAA AGCCTTAGTA GTCAGAAAAG CCTTGTGCGA AAAAGTTTAA
215761 ACCTTTAAGA ATTGCACACA TGGAAAAAGA TCAAGTAAGC TATATATACA CCATCTTAGC
215821 AATGATTTTG AAGTGAGAAT TAAGGCTACC ACAGCTCCAG GTGGTAAGGA GAGAAATCAG
215881 GCTGGAAGAG TTTGAAGTTT CTGTATTATT CTAAGCTCTT TACTATTCTA TTATGAGCTC
215941 ATTAATTCTC ACAACAACCC TCTCATATAA GTACCATTTT AAATTCCTAT TTTACAGAGA
216001 AGGGAGTTAA GGAAGGTGGA GATTAAGAAA ATTGCCCAA TACAAATAGC CAGCAGGTGG
216061 TAGGTCTGAG ATTTAAGCCC ATGCAGATT TAGCCCCAGA GCAGACATTC TCAATCACTA
216121 TGCTAGACTG CCTTTCATG GTATGTGATC CTACTCAGGC CTCTACAGCT TTATCATTTG
216181 TGTTCTCCCC AGCCTGTCGT GCTGAGAGTA TATACTCGAA GAGCAGAACT AAAATTCAT
216241 CCAGCTTCTC ACTCCTAGGT CCACTACACA GCTGCATCCT GCAGACTTTT ACCTCAAGCA
216301 ACCCTCCTGC GTTCTTGCTT CCTTCCATCA TAGTTGTAAC CATCTCCTCT ATTTGCAAAAT
216361 ACTATCTGCT GATCTCTCTC TTCTAGACTG GTTCTTTTCA ACCTTCTTCC CACCAAAACC
216421 AAGTTAGCTT GCTAAAATAA AGATGGCGCA TTTTACTTCA CCCGCTTGAG AATTTTCAAT
216481 GTGTTCCCTC ATGCTTACAG AGTAAAGCCT GACCTCTTTA TTGCATGAAT ACAAAGTTC
216541 TTAGCCATCT GGCCCCAACC TTGTTCCACT CAACTCCCCT GTGCAAGCAT GGCTCCAGTG
216601 GCACTGGACA TTGGCTGCTC TCCACATAGA TCTGCACTGC ACTTCCCTCT GGCTCTGCTC
216661 CCGTTAGTTT ATATGCCTGG AAAGTTCTTT GCCCCTGTTT CTTGTGCCAA AATTCATCT
216721 ATCCTATTGC ATAGCTTATG TAAAACTTC CTAACCTTTT TTTTTTTTTT TTTTTTTTTT
216781 TTTTTTTTTT TTTTTTGAGA CGGTGTCTCA CTCTCCGCC CAGGCCGGAC TGCAGTAGCG
216841 CTATCTCGGC TCACTGCAAG CTCGCTCC CCGGTTTACG CCATTTTCTT GCCTCAGCCT
216901 CCCGAGTAGC TGGGACTACA GGCGCTGCC ACCATGACCG GCTAATTTTT TGTATTTTAA
216961 GTAGAGACGG GGTTCGAAGC CAGGATGGTC TCAATCTCCT GACCTCGTGA TCCGCCCCGC
217021 TCGGCCTCCC AAAGTGCTGG GATTACAGGC GTGAGCCACC GTGCCCCGCC AAAACTTCTT

Figure 9 (Page 67 of 74)

SUBSTITUTE SHEET (RULE 26)

156/162

217081 AAATCTTATA ATTATTATCA ATTTATCCTC AGATATACTT CCACGTACAT TGTAGTTTTA
217141 TTATATTTAT ATTTTACATC TTTTTTTTCA AATTGCAGTT TGGGACCCAT TAGTGAGTCA
217201 TAAATCCAT TGAGCGGGT AAAATCATTA TTTTAAAAA TGAGTAGAAT AGAATAGAAA
217261 TTGTTGGAGT GCATTGGACA TGGTAAAGT AAATATCGAT TCATGAAACC ATCGTTTGAG
217321 GCATATGTGT GTGGTTGTAT GTACAAGTGT TTATGCATAT TGGTGTGTGT GTTATGTTAC
217381 CCTGTAAAAT GCATTCTTA CTATAGGTCT CTGTGAAATA TGTGTCTTGT TGTTTTTTAA
217441 TGTAGACTTC CAAAGCCTAC ATGGCATTTT ACTAGTGACA ATCAATTTTA TTCACATTTT
217501 TCTCTCCAAT TGGACCAGAA GCTCTTTGAG GGCAGGGGCT GTATCTTACC GATTTTGTGA
217561 AGTCTTTTCAT TTCCTGCCCC TAGCCTCATA TTAGATCATG CAAGAATGCA ACTGTAATCA
217621 CAAGAAAATG CTAATGGGCT GTGATAGCAG AGAGTTACTG TGACAACTA AGGGATTAG
217681 ATTTGGTCAC ATTGGTGTG AGGAGCCATT GAAGAATCAG AGAGTGTGTT ACTATTATTT
217741 GTTAATTTTA ATTATATCAT ATTACTTTAC TGGGGAAAAT CTGTGAGCTA TTTTAGAAAT
217801 AAATACTCTC ATTGCCCAAT AATCTAAGT CTGCCACCTC ACTGTGGGA CATTGTTTAG
217861 GGAGGCCACG AAGTCTCAGC CTTTGATATT TTCATAAGTG TTTTCTCCC TTTTCTTTT
217921 AGGGTCAGCA TTTGGATCCT TCATCATCCT CTGTGTGGGG GGAATAATCT CACAGGCCTT
217981 GAGCTGGCCT TTTATCTTCT ACATCTTTGG TGAGTCACTT TCTCTTAAAT CCTAATGCCT
218041 CCATTTCTTG AGCATCCATT TTGGCACCTA CACCACCAC ATTCTTCTTA TATGAAAGAA
218101 AATGTCCTTT ATCAAATGGA AGATGATAAA AAATGTCAAC GGTGGGTATC ATTTTAAATC
218161 TAGTCACACA ACCTGATTAA CACCTTCTTG GTGGTTCTGG GAAGCCACAC GCAAAAGGTA
218221 GAGGAGTTGA CTATTCACAT GGCACCCACC GACTTGTGAT GCAGTCTTGT CCTTCCATAT
218281 CAAGCACCTT CTGAGAATC TCTACCACCA CATCTGAAGT GCCTGCTATA TGCAGTTAAG
218341 ATGTCAAAGA TAGTGAAGTA CATTTTCAAT GTGTCTTCAT ATTTCAATTAT AATTATTATT
218401 TCTGTCCAAG ATGCCTTTCA CCTGTTCTCT ACCAAGTTAA TCTTGCAAAG TTCAATTCAA
218461 ATGTTCCCTT CCCCATGGGC CCTTCCAGGG CTTACCCTGT CAGATTCTGG CATTCTCTCC
218521 TTTATGATAT TTCCTCTCTA GGTATGTTG GTGTGTAATT ATTTATTTCT CCTTTTCTTT
218581 CCACTAGACT GTGAAATGCT TGAGGCAAGG AATCCATTCT ATGTTTTCAT CATTGGGTG
218641 TCATCATGGT GCCTGATTTT TAGCTTTAAA ATAAAAGAAT CAGTGAATCC AGTAATTAGA
218701 GGGGATTTAA AGAAAAC TAG TCCCTCAGAAT CTTTTAACAT AGAATGTTCT TCAATAAGG
218761 AATTCCAATA ATAAGACAAT TTTCTACACT TGATTTTGT TTTATAGCCA AATGGTGTCA
218821 TTAAATATAG TCCTGGCCTG AATGGCTTTC TCATTAATGA TGCTAATTAT TTTGGTTTGT
218881 ACATGTTAAC CAGGTATTGT ACAAAAATAT TTCTTTTGGG AATCCATAAT GGATGTATGG
218941 CTTGAATACA AATAATACTG TCTCTGTAA GTGCATTGGA AATTTTCCC TGCCACATGA
219001 TTTTCATGGAA GGTGTTTTCG TGTATGTATG ACTGCAAACC TGACTATTCA GATCTTCCGC
219061 AACAAGACAA CTTATGTGTG CATTAAGAAG TTGCTGCCTA AAATACATAA CACTGTAAATC
219121 ATTGGAGACT TTAAAGTAAT TAATCAGCTA TGCAATGCCA CGCTCCTGTT ATCTCCAGAG
219181 GGCTCTGACA TTGACAAATG GTGGCTTTCT ATTTGAGACG TAATATCTAA AAAGCTTTAA
219241 CAGGTTTGTGA GAAGGATTGA AAGAAAGAAT GGGAACATTT AGGTCCTTAT GGTAGAATAA
219301 GCATTAATTG ATTAGTGTGT AGAAGGGAGA GGCATGCCAC TTCAGAGGAA ACTTCCTTCC
219361 CCCAGTAAAC AAATCTACCT AAAAATAAT TTTATCCCTT CTTCCAGGT AGCACTGGCT
219421 GTGTCTGCTG TCTCCTATGG TTCACAGTGA TTTATGATGA CCCCATGCAT CACCCGTGCA
219481 TAAGTTTAG GGAAAAGGAG CACATCCTGT CCTCACTGGC TCAACAGGTA CAGTGCACAC
219541 CTTGTACCTG TGGCCCATGC AGAGTCTCT AGGGCAGGGT GTGGATCTCC TCTGAGAGGC
219601 ACCATCTTGG CTGCTCTAAT ACTCATGCTG ATTAGATCTT TCTTTTTCAG CCAGTTCTCC
219661 TGGACGAGCT GTCCCCATAA AGGCGATGGT CACATGCCTA CCACTTTGGG CCATTTTCTT
219721 GGGTTTTTTC AGCCATTTCT GGTTATGCAC CATCATCCTA ACATACCTAC CAACGTATAT
219781 CAGTACTCTG CTCCATGTTA ACATCAGAGA TGTGAGTTTA CTTCTTATAC TTCTACGAAA
219841 ATGATAATGG TAATAAGGAG AAACAGTTCT GTGTTACCTA TTACATTCTG GCTTTACATA
219901 TAACCATTAA TTTAACCTT ACAATGACCT TGAGAGAGGC ATTGTTATAA TTCCCTTTTC
219961 ACAGATGTGG AAACAGGACA CTTAGAGGTG AGATAACTTG CCCCAGGTTG CACAATACTA
220021 AGTGATAGAG CTGCTGCAGC ATCCATATTC TTAACCACTA TGCTATACTA CCACACCAGC
220081 TGATTCCAAA GCTTCTTTTA GAAATAATAT TGCTGGGCCA GGCATGGTGG CTCATGCTCG
220141 TAATTCCAGC ACTTTGGGAG GCCGAGGCAG GCAGATCATG AGGTCAGGAA TGCAAGACCA
220201 GCCTGACCAA TATGGTTTAC TAAATATCAT CTACTAAAAA TACAAAAATT AGCCAGGTGT
220261 GGTGGCAGGC ACCTGTAATC CCAGCTATTC AGGAGGCTGA GACAGGAGAA TCGCTTGAAC

Figure 9 (Page 68 of 74)

SUBSTITUTE SHEET (RULE 26)

157/162

220321 CCAGGAGGTG GAGGTTGCAT TGAGCCAAGA TCATGCCACT GCACTCCAGC CTGGGCGACA
220381 GAGTAAGACT CCGTTTCAAA AACAAAAAAC CCAAGAAATT AATATTGCTT TTATCTGGAG
220441 CCCAGAGTGA TGCAGCTTCT GGCCCTCTTA TCTGAGACAG TGTTCTTTTA GTGTGAAAAA
220501 GGATGCTAAT TTTCCCCCAA ACAACCCACA GTATCATGGG GGTAAGTTAA TGGCTGGTCT
220561 GTGTAACCTGA CAAATTTTGG TGCTAACGTA TCTCTATAAC TACTCTGTAT AAACCTCCTT
220621 CCTTCAGAGT GGAGTTCCTGT CCTCCCTGCC TTTTATTGCT GCTGCAAGCT GTACAATTTT
220681 AGGAGGTCAG CTGGCAGATT TCCTTTTGTC CAGGAATCTT CTCAGATTGA TCACTGTGCG
220741 AAAGTCTTTT TCATCTCTTG GTAAGGATAA GCGTGTGGGC CCATTTAACC AATCCCTTTT
220801 CTGCACATGG TCTCAGAGGG TTCCCTGACA GCATGTCCTC ATTGCCCAGG GCTCCTCCTT
220861 CCATCAATAT GTGCTGTGGC CCTGCCCTTT GTGGCCTCCA GTTACGTGAT AACCATTTAT
220921 TTGCTGATAC TTATTCCTGG GACCAGTAAC CTATGTGACT CAGGGTTTAT CATCAACACC
220981 TTAGATATCG CCCCCAGGTA AGAGCTCTAC CTGTTTTTTC CCCTCCTCCA GACCCCTCCA
221041 GAGGTGTTAG ACCTCAGTGG TCGCCGTGAA ACTCTTTAAT GTTACTGACA TTGCACTAAT
221101 GGCAGAATGA CAAATAACTA CAAATATCTG TCTGTGGCCA TTTTITAGAAC AACAAATGTG
221161 GCATTTTTAG AACACAATTT TCCATCTTGG GCCAGTAATC ATTTTGACAA AAACCTTCCC
221221 AAGCTTCCCT AACAGAGATT GAACTGTGTA TGCTGGGAAA AGGCCACAC ACAGGTGATT
221281 TGGAAAAGTT TCCATGGTGT TGTTTATATT AGCTACCACA TATATATATA TATATATATA
221341 TATATATATA TATATATATA TATATATATA TACAGTCACA ATAAGCCAGC TCCTGTGCCA
221401 AGACTTGCCA TATATCAACA CATCTAATCC TCACAGTTAT ATTAGGTAGG CCCTATTGTT
221461 ATCCCCATTT TATAAGGGAG AAGGCTGAGG CACAAGGAGG TTAATGGTG TGACTATGGT
221521 CACATAAAGG CAGAGCCAGG ATTTGGACTG GGGGAGTCTG GCTTTGGAGT CTGTGTCCTG
221581 CCCGTTGCAC AAACCTGGCTT CTACACTGAG CAGCCAGGGT AAAGAAACGT GGTTCCCAAG
221641 GAGACTGCAT TGCTCCCTGG TTATTGACTT GGTAGATTGG TAATTTTCCG TTTGGCAAAAT
221701 AGACATTGCC CTGAATGTCT TTAGGTGAAT GAAAAACTGC ATTAAGCAAA ATGACTTTGC
221761 CATTAGAGCT GAATTGCATT AAAGTTGAGT TGCTGCAGAA GCTGTAGGTG GCTTCTCTATA
221821 TAAAATCATT TATAAATCA TCTTCCATA GATATGCAAG TTTCTCATG GGAATCTCAA
221881 GGGGATTTGG GCTCATCGCA GGAATCTCT CTTCCTACTG CACTGGATTC CTCATCAGTC
221941 AGGTTGGGTC AGTTTATTGA ACATCTTCAA GTGGCAGGTA TTGTTTGGT TGTTGGAGAT
222001 ACACACGGTG CTCTAAAGAT CTGGATGGCA ACACAATTAC TCTATTTACA TGAGCCTCTA
222061 AATCAGACTC TGGTAGGTCA GATTTCCAG AGGAAGAAAA ATATAAGCTT ATTTCTCAA
222121 GATGAATAGA TGTTAGATTG ATTAATATGA GCTGTTCCGG TGCAGAAGAC AGCACGTATG
222181 ACTTCCTAGA GGTACATGAG CATGAAACAG TTCTTAGTTA TGACCAGAAT GAAAGACACA
222241 TGTCAAGGAA TAGCAAGAGA CGAAGACAGA GGGGCAAAAG AAGATCATGA AGAATATGTT
222301 CAGACTAATC CAATTTTTAA AAAATCACA AAGGGAACA AAGTGTCTTA GGCCAGTTTA
222361 AAGATAATT AATGTCTGGA AACAGATCGG CTGTGAGACA TTGCAAGGAG GCTTGGCTCGG
222421 TGTTTGGAAA TGCAGGCTCA TGAGGAAGAT GAAAAGACAG ACCCAGGCAG GGATGGAAGG
222481 ACTGACTAGA ACCAATTAC AAAGAGAAGT TTTGTTTTTA CTACATTTCT ATGTGATCAA
222541 GTTCCCAGGT TAATATTTGA CTAACTGCT AGGAATCCAC TGTGACTATA ATGCTGGAAA
222601 TGACTTAGTA GGGCTTTCTG AGGAGGGTCA CACAGAAGAC CAAAGAGAAC TCATGTTGAA
222661 TTGAGATGGG TTATAGTGAT AGTTGTCAAC AGCCAATACA GAAACAAAAA AAAACAAAAA
222721 AAACAGCAAC AACACAACA ACAAAAAAAA AAAACAGAGA AGACACAAAC ACAATGCCAC
222781 AATGCCATTT TAGGCATAAT TTAAATGAG TAATATTATA TGTTGAAATC CAAATTTTCA
222841 GAAAAACATT AGTGTATTTT ATTTTGTGTT AAAGAAATAA CCATCTCAAC TCAGAACCCC
222901 ATGTGCATTT TGGCCATTTT GTTCCCAATA GTTTCATAAA CTTTCTTAAG TAACACTGTC
222961 ACATTGTTCC TTATATTCCT TGTGATCAAC ATTGCAATAC ACAACTGGGA GGGCTACTAG
223021 AACTGGTGTA GAAGGAACTT GTGAGATTGA TCATTTTCTC TGTTTTTTAT ATCTAGGATT
223081 TTGAGTCTGG TTGGAGGAAT GTCTTTTCTC TGTCTGCTGC AGTCAACATG TTTGGCCTGG
223141 TCTTTTACCT CACGTTTGGA CAAGCAGAAC TTCAAGACTG GGCCAAAGAG AGGACCCTTA
223201 CCCGCTCTG AGGACATAAA GTTACAAACT TAAATGTGGT ACTGAGCATG AACTTTTTAA
223261 ACATTTTTTA CTCTCTCCA TATTCTGAC CATAGACTCA GCAGTTCTTA ACTCTGGCTG
223321 TGTGTTAGTC TTCCCTGGGG AGCCTTTATA AGACACTGAT ACTTGGGACC CACTCCAGAG
223381 ATTCTGAATG AATTGGTCTG GGGTGGAACC CAGATACTAC TAATTTTTAG ATACTCTTAA
223441 GAGGTTTCTA GCATGCGCCC GGGTTGACA ACAGCTGGAC AAACCTGAAA AGTCAATTCA
223501 TGTGGCCTTT GAATTTTCTT CATTGGAAG TACTAAATAA ATAAAAATTC ATGTGAAAAAT

Figure 9 (Page 69 of 74)

SUBSTITUTE SHEET (RULE 26)

158/162

223561 GATCACTGAT AAATATCTTC ATGGTGGGGC AGGTTATTGG ATGCAGAGAA GATCTGCTCG
223621 GAATTGTAGC CATATGTTAC AGATCTCAGC ACCGATCAGA ACTGTAAAGC TATAATCCCC
223681 AGAATTAAAG TTTTATTAT TTTTATACA TTGTAAACA TAGACGTTTA TTTATGTGAT
223741 TAAATTCTAT TAAAATTTAC ATGCTAAAAT AAAATAGACC ATTTTCAAAT TATTTAGATC
223801 CAGATATTTC CATCAGATTA AACAGATATT TATTTATCCT AGCCCAATTG CAAGAGATTA
223861 ATGATGAGAA AATGACCAAT ACAAGATTAA ATAAATGAGG TTAACCTAGA AATCAAGGAC
223921 AGAGAAGATA GAACTGGAAA GCTTGTATTG TGAGAAGAAT GAATGTGAAG GAAGGCAATG
223981 TAGACACTTC CAGAAGGGAT AGCAATATAG TTTAGACCAT ATAATGAAAA TTGGAGAGAG
224041 ATGACAGAGA CACTTTCAAG TGAAATGACA ATTTATATGG GGGAGAAAAA TATTGAAGAC
224101 ATAACAAGAT GAGAAAAGGC ATAGAAATGT ATCACATACA AGGCATAGAA GTGTATCACA
224161 TACAAGAGAA GTTCCTTTTG AGCGTAGAAA AAGATAATTT AACCTTCTTC ATATTTTCT
224221 TACTTTCCCA AGATACTCAG ATAGGCAGCG TCAACTCTAA CAGGAATTAA TTTGGCTCCT
224281 AACACTTAAG ACATATCCTT TAGTTTGTCT CCTCACACAG AACTGATTCT GGTTTGGCCA
224341 CAACATGTCT AGAGAAGAAG TTCCCACCAT ATTTTAAATC CTATTAAAAA ACTGCTTGGA
224401 CAAGAACCTT GGGCTAATTC AGCAGATGAA GAGAATCTCC TAATGCAAT CAATGGGTAT
224461 TTTTGAGCAA GTTTTTCAGA AAAACAGAGT GTCAGGCCCT GAGGGTGGTA CTAAGATGAG
224521 AACATTGATT TTGCCTTCAT GATATTGACA ACACAAAGAG GAAAGGGGGT TTGCAGAAAA
224581 CTAAAAGAAG AAGTAGAAGA AAAAAGAAAG ACATAGTATA ATAGGTAGTC AAATTATGTA
224641 CAGAAAAAAG AGGAAAAAAA ACCAAAAAAG GGTGGGGGAC AGACAACCCA ACTAAAAAAT
224701 GGGCCAATGA CTTGAACAGG GACTTCATAA AAGAGAAAAT GTAAGTGGCT CCTTAACATA
224761 TAAAAAGATG TTCAACTTCA TTAGTCATTA CAGAAATGAA AATCAAAACT ACAATGAAAT
224821 ACCACTATAA AATTAECTAA TGGATAAAT GAAAGGAGAT GGAAACAAA ATGTTGCCAG
224881 ACATGTGAG CAACTGGAAC TTTTACATG TACGAATGTG AACTTTGGAA AGCTGCTCGG
224941 CAATATCTCC TAAAGCTAAA TGTACAATTC CAGTGACTCA GACATTTTAC TTAGAAATGC
225001 ACATATACAT CCATAAAACA TGTACAACAA TGTTTATAGG AGCACTATCT GTAATAGCCT
225061 GAACAGGAAG TTGTCTGTTA AAAAAAGAAT GAGTAAATAA ACCACGGTCT ATTTGTATAG
225121 CAATGAGAAT TAACAGACCC CAATATATAA TAGATGAATG GGTCTCATAA GCACAATATT
225181 GATTAAAGGA AGACAAAACG CACATTCTTT TAAAGGTTTA TAAAATACTT TTTAAAAACA
225241 GCTACAACCA ATCCGTCTCG TTAATAATCA GTGAGCGATT TCCCTTGTGC AGGGATGGGG
225301 GTTGTGGCTG GATGGATGGT ACTTAAGAAG TGCTCCTGGG GTACTAGAAA TATTTTATTT
225361 CTTGACTTGG ATGTGTGTTT ACTTTGTGAA TATTGTACAT TTATGATTG TGCACGTTTA
225421 TGAATGTAGA AAATAAAACA GAAAGCAAT TCAAAGTATC ATCCTTTTGA GAGCTTCTGC
225481 TCTGACTTCG TTTTGACCAA TGGAGCAGTT GGAAGGGGT CTTGGTCCTT CGGTCTTTG
225541 CTTTTTTTTT TTTTTTTTTT TTTTAGACAG AGTCTCACTC TGTCGCCCCG GTGGAGTGC
225601 AGTGGCTCGA TCTTAGCTCA CTGAAAGCTT TGCCCTCCCG GTTCATGCCA TTCTCCTGCC
225661 TCAGCTCCC CAGTAGCTGG GACTACAGGC ACCTGCCACC ATGCCCGGCT AATTTTTTGT
225721 ATTTTTTAGT AGAGACGGGG TTTCACCATG TTAGCCAGGA TGGTCTCGAT CTCCTGACCT
225781 CGTGATCCGC CCACCTGAGC CTCCCAAGT GCTGGGATTA CAGGTGTGAG CCACCGCGCC
225841 CGGCCCTGG TCCTCTGCTT TCATGTTCTT CTTGGTCCTG TTCCTCTCC TCTTTTGTG
225901 GAACTTCCAG TATCAGAGCA GGAAGGAAG CAATGGGTCA ATCGATGCTG TCAGCTTTTG
225961 GATCAAACTG CAAGTTCTCA AACAGCAAAA TTAATGAGCT CAGGCTTTGA AGAAACCATG
226021 ACCCTGAAAG CATCAGTTGC TTCCAATTGC ATCAGTTGCC ACGGTGATA AGAACAATGA
226081 TGACTCAGAA TGCCTAGGTT TTCCAGCAG CTTCTCTGAG GTTTTCCAG CAGCTTCTCT
226141 GATTGATTCC TGACAGATGA CTTCCGTGTG TCAGACTTTC AGGGTATCTT TCCTTATGTG
226201 ATGGTTTGAG GAAGAGTTAC CATTACATT CCTAATGGCT TCAGAATAGA TGCAATTGTG
226261 AACTGATAGG AAACATTTCT AATTCACTCT CCCTCCCAT CCCTAAAGGA TTGTTTCTAA
226321 CAATAGTCAT GAAATTAAT TCACCTTCT CAAATAGTTT ATTGTCATCT ACCTAATGAT
226381 GAGATGACTT ACTTTTCTC CTTGACTGTT AAATATTATG AATTATATTA ATGTATTTCT
226441 TAATGTTGAG CTTTCCCTG AATATTCTTT TGATGTACGA CAGAAATTTGA TTCACTAATA
226501 GTTATTTAG GACTTTGGCT GATGTACTGA TATATGAGAT TGGCTCTGTA TGCATACATG
226561 TGTTTTGTGT ATCTTTTTTG TGTCTGATA TGGAGCTTAT GCTGATTTCA AAAACAAGAA
226621 AGGAGAACTT TCCTTTTTCC CCATTACTCT GAAAAAGATT GACTAGAATG GAATTTTTAT
226681 AATGCTGTT GTTATTTGAA AGCTTGAAAG CATTGGTTTG TAAAAATCAT GCAGGCTGAA
226741 AGCCATTTTG AGGAGACTTT GATAACTTTC TCAATTCCT TCAGTACTG GTCTTTTAAG

Figure 9 (Page 70 of 74)

SUBSTITUTE SHEET (RULE 26)

159/162

226801 GGGTTTTATA TTTTCTTTG ATCAATTTTG ACCATTTATG TTATCTTGGA GGATCATCTA
 226861 TTTTACACAC TATTTAAAGT ATATTTGCAA AAATTCAACT GTTTTATCAG GCTATCTTTT
 226921 TAATAATATA TTCATTTTAT CTATATCTGA GGTTTGTAGT TCTTGTACT TCTGACCCAA
 226981 TTGCATGTGT GCTTCTTTT TCCTTCATTA GACTACTTAG TCATTTACTA ATTTTAAGAA
 227041 TAGCTTGTCT TTTATTTATT TACTTATTTA TTTTGTAGAC GGAGTCTCAC TCTGTACCCC
 227101 AGGCTGGAGT GCAGTGGCGC GATCTCGGCT CACTGCAACC TCCGCTCCC GGGTTCAAGT
 227161 GATTCTCCTG CCTCAGACTC CCGAGTAGCT GGGATTACAG TCATGCACCA CCATGTCTGG
 227221 CTAATTTCTG TATTTTAAAT AGAGATGGGG TTTTGCCATG TTGGCCAAGC TGGTCTCAAA
 227281 CTCCTGACCT TAGATGATCT ACCCACCTTG GCCTCCCAA GTGCTGGGAT TACAGGCATG
 227341 AGCCACTGCG CCCAGCCCTG CTGTCTTTT TATTTTATAT TTGATTAGCT TTATCTTTTA
 227401 TCAAGCTTAT GTCCTATTTT CCTTTGCTTT ACTTCATATA AATTTTGTTT TGGATAGTTT
 227461 ATTTATTTTT CATTTAATTA TGAAACAGGT TAAAGCTTAG AGGAAAATTG CTCCTCTAAG
 227521 TCCACTTTTG TGGGCAGATT ACATTTTGCT GTGTTGTGCT CCCAAATTCA TTGTTCTTTT
 227581 AATGCTTTAT TTCTCAAGTT AATAACCTAT ATAGTAAAA AGTGGCTGTT GACTCTCAGC
 227641 TTTTTTTTTT TTTTTTTTTT TTTTTTTGTA GATACAGGGA TCTTGCTGTG TTGCTCAGGC
 227701 TGGTCTGAAA CTCCTGGCTT CAAGGGATCC TCCTGCCTTG GTCTCACAAA ATGCTGGGAT
 227761 GACAGACATG AGACACCATG CCCAGCCATG TCTCTCTCCT TATATATAAT AAGAAAACAG
 227821 ACACACTGAG GCATCCTATC ATCTCACTCT TGGTTTCACT ACTGTTCTCT GGAAGTTTTG
 227881 CTCTGACCTT TTGCAGTTAA TGTATTAATT TTGCATTGAG TAGTTTCCAT AGAAGAATTA
 227941 TAGCATTGTC ATTCTGTTGG GTATTATACT TTTCACTGTT ATTTGAACAT AATTTGAGGG
 228001 CTGAAACCAA GATGAGGCAA GTGAGGTGCC CAGGAAGCAA TATTTAAGGA GGCATCCTTT
 228061 CTTAGGCTCA TGCAAGAACA GAATTGGCAC ATGAGAGTGA GTGCCTCCTT AATTTTGAGT
 228121 GCTGGACACT TCTTGCTCAC TTAGCATACC CCTGGACAAT GAAGTGTTT TTGTTTGTGTT
 228181 TTTTCATGTC CATCCTTTAT CCTTCTTCAT CTCAAAACAT TTCAATGGAG TATTTTTTTG
 228241 GAGCAGTACT TGGATGAGCC TCTGAGTCCC ACAGTAGCTG AGAATTTATT TCATAGTACT
 228301 CTTTATGATC ACTGTGGAGC CTTAAACAT TGTAATATTA ACTTAGCTGG GAACAGAAAT
 228361 TTTGTCCAC AATTTGTCTT ATTCAGAACA GTATTGACTT CCTGCTAGTC TCTTCTGATG
 228421 TCCAATATGA GGAAGTCTAG TTAGCCAGCT ACTTTTGTGA GGAGAGCTAT GTTTAGGCTA
 228481 GGTGCTATAG GATTCTCTTT ATCCTGGAAT TCCTTCACCA AGATGTGCCA AGGTGTTAAT
 228541 CATTTCTCT TGCTTTTGG CTGGTGGTCT TAGAGTTTCC TTCGATTTG TTTTATTTAG
 228601 TGATTGTCTT CAATTTGTTT TCTTTACTAA GAATCTCTCT TCTATTTATC TGTATGGTAA
 228661 AACCTTGTTG CCCATCTTTC TGGTTTCTGC TGACTTTTCT TTTTGGACCT TTTACTTTGC
 228721 TTTCTCCATG GACTTTTGG TAGTGGAGGC AGGCAAACAC TTTCCAAAGT CTTTCTCAAT
 228781 TTCCATCAAT TTCAACTTAT TTCCTAAAT TGCCCTCAGAA TGTGCTATG TCCACAATAT
 228841 CCTCTCTTCC ACTTAGAAA GGAAGGCAT CCACACTTTA TTTAGGTGCA ATGCCTGAAG
 228901 TGTAACACTT TCTGGTTGT CAACAAAGGA GTACTTCCAA ATATTGGTTT GGGGATAACC
 228961 TGCTAATGAT TAACACATTC ACCTTGGCTC TTGGTTTGCC TGCTCCCTCT TCTTTTATCT
 229021 GCTGTGTGTA TTTTTTTTAA TCACTGAGAA TATGCACAGT ATTGTATGTT TTATTATAAG
 229081 AGAGGACTGG CCAGAGTGGG AATGTTCTGA ATTCAGAATA ACTGAAGCAG TACAGGATAG
 229141 GAACTCATTC TTCAAATGA AGCTGGCATA TTTTCCAGA GCACCAAATT TCAATATATA
 229201 TTTAAAAAAC TTGATATGAA TGATACAATA AAGTGGTTAG AACTTTTATT AAAATAAACT
 229261 TATGTCATGA AATACTTATT CTAATTATAG TCACTCTTCA TCTTATTTCA TCTTATAACA
 229321 TGTTTAATGT TTTCTTTTAT TTACAAAACA ATTTATTTTT TGATGAAAAG TTTTAGAAAT
 229381 CAAGTTAAAA ATATTCAAAG GAATGCCTAA AGTTTTCAA ATTCTTTTAC ATGTTGTACA
 229441 ATCAAAAGAG TCTGAAGACC ATTAGCTAT CCAAATTGTT TATTTTAAAG CAGTATCCCT
 229501 TCTAATATTT ACTATTTATA ATCCTTAAAA ATTTGCCTTA GCACAGGAGA ATTGCTTGAA
 229561 CCCAGGAGAC GGAGGTTGCA GTGAGCCAAC ACAGTGCCAC TGCCCTCCAG CCTCGGCGAC
 229621 AGAGTGAGAC TCTGTCTCAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGCC AAAAACAAAAT
 229681 AAACAAACAA AAAAATCCGC CTTAACATTA TTTGTTTATT AAAAATTTT TTTAATACTA
 229741 CTAGTTTCCC TTCTCTCTCA GCCCATTTGTC ATATTTTGAT TTTTATCACT TGCTTTGTAG
 229801 GACATATGAG GTTTTTGTTT TTTTTTTTTT TTGGAGATGC AGTCTCCCTC TGTGCCCCGT
 229861 GCTGGAGTGC AATGGCGCAA TCTGGCTCA CTGCAACCTC TGCCCTCTGG GTTCAAGCAA
 229921 TTCTCCTGCC TCAGCCTTCC AAGTAGCTGG GATTACAGGC ACCCACTACC ACGCCTGGCT
 229981 AATTTTTGTA TTTCTGGTAG AGACGGGGTT TCACCATGTT GGCCAGGCTG GTCTCGAACT

Figure 9 (Page 71 of 74)

SUBSTITUTE SHEET (RULE 26)

160/162

230041 CCTGACCTCA AGTGATCCAC AATCCTTGGC CTCCCAAAGT GCTATGATTA CAAGCATGAG
230101 CCACCTGCCC AGCCAGAATA TATGTTTCATT TTGAGTCCTT TAACAAAGTC ATAAGAATTT
230161 TAGGAATTCA GTTACTTTCT TGAGAAAATC TCTGAAAAGA TGCCAATAAT TTGTAGCCAA
230221 TTATATTGAT TTCTCTTTT CATATTGAGA ATTGTTTTTT AAAAAGTTTG TATGTGTGAA
230281 GATTTTGTGA CTGTAGTTAA AGAAACCACC TGTGTGTTGG TTAAGCCATA AGTACATGTA
230341 TTCAAATAAA TTGAGGTGGG GTTACTCTGA GAATCAAAGG AAAACCTGAA GAAACAGGCA
230401 GCCTCAAAAG GTCTTAGCTG TAGCAACTTG CTCCATTGTT GAAATAAATA GGCTTGAAC
230461 TGTATTTTCC CTCTACTCAA CATTTAAGGT CTCAGAAGAT AATATAATTG GTGAAATTTA
230521 AGTAAAGTGC TCACTCTTTT GCTTTAACAA ACCCTAGAGA GCTGGTAGGC AGAGCCTCAA
230581 CAGACCGTTT TAGCTTCCAA AGGGAGTTCA GGACACCATG ATTCACGACC ACAATACATC
230641 ACACATAATT GAGAAAAGAT AGTTCCACCA AATAAAGTTG AAATGCTGAC AAGAAGGGGT
230701 AAGAAATCTT GGAAATAGGT TTATATAAAA TTTATTTTTT CCTTTTTTAT TGTATGGAA
230761 TAGGACCAGT TCTACTTAAG CCACCATTG GCCAAAATAA AGTGAGAATC GTTCTTTTG
230821 GGGACTCCTC TTTGTAGCTC CAAGTGCCAC TAACAATTCT TAGGACCTGA GCTATAAGCC
230881 AGGTGATTTT AGTTAATATG ATCAATTATT TCATTTAAAT GGCTCTAATG TGCAGAGGGA
230941 ACGGAGCCCA TCAGCATTCC CTGCAGGGAA CTGCAGTGGC TTTTATCAAC TTGAACAGCT
231001 AGCTTTCAAC TGTTTTGAAG TCACTTTCAG GGTGGTCATG TAGTTGCTTT TTTGAAATCA
231061 GAAGATGATT CTGCCTCTTT TAATATGTA CTCCTCAGAT TCAGAAAGTG CTCGCTAGTC
231121 TTAAGAGTGA ATTACCTCA GTGGTCCAGC GCTTATGAAC CCACATCTAA CCCTATCCCC
231181 TGGGGGAAC ATCAGAGAAA TTGGTGCCAT GGACATAAGA GGAAGGCACA GTGAAGCAGA
231241 GAGCCCCGCA TGATGAAAT CAGTGACAG CATCATTATT TACAACCTTG TAATCACCCA
231301 GGAGCATGAA AATCCAGGCC AATCTGGCAC CATGAGCTCT AATTTTGTG GGAGTCTTG
231361 GAACCGATTG TGATGAATGA CTGTTTAGCC ATTTTAGAGT GTGGCATACTG TGGCTGCTGG
231421 CATACAGAGG TTGGATGTAA ACGGGCCTTT GCCCTCTCTT ATGAACATAG ACAGGAACTA
231481 AACTGTGTCA CATAGGTTCC AAATGGTGGC CTGAATACTA TTTACAATA AGGTACAATG
231541 AAATTGAGTA AGTCTTTTCC TCTTTTGCAG ATACCATCAT TATTCATATA TTCTTCAAA
231601 GTTAACTATT TGTATTTGGT AATTTTAAAT AGAAATGTAA TAATTGCTTC TCAAGTTTAG
231661 TCTTTAGTCT TAAGGTTGAT GCTCTCCATG TCCTTCCAAA AAAAGGTATG TTGCTTTTAT
231721 TATATCCTCG CCTTCAGATG GGATTATTCC ATTTTGTCTT TGTTAATAT ATACTTTGAG
231781 CCACTTTTTT TGTGGCTCTG GGTGAGATGC TATAGGTACA ATGACAAGTG ATACGTGTGT
231841 TGTCCCTGTC ACAAAGTGG ATAGCCTAAG TGGTGAATTT TACCTCCACT CCAAATATAT
231901 GTATCACACA CCAGCCGTAT GCCAGGCACC ACTCTAGGTG CTAGGGATAC AGCAGTAAAC
231961 AGACAAATGC AACCCTGCC CATGTGAAAG AGAATAAGAC AATAAATAAG TAAAGTGCAT
232021 GTTATATGGA GGTGGCAAAT GCTAAAAAGA AAAATTAAGC AGGCAAGAGG ACTCATTGAA
232081 AAGATGACAT TTGGGTAAAA GCCCATGTAT ATATGTTCTA TTGGTTTTAT TTCTCTGGAG
232141 AGCCCTGACT AATACACAAT GACTTTGAGA AGTTACTGGC TTTTGATTTA TCACACTATT
232201 CGGAGTGCTG AGAGCCTTCT TAGTGTGTAT TCAGTGTGTT AAGAGAGCTT GTGGATGAAT
232261 AATAAATAGG ACAAATTTA TCCAACTTA AGCCTTGCTT TAGGTAAAAG GGCTCCTCTT
232321 ACAAGGTAGA AGGTTATTAT TTGACATTTA AATCCAACTG AAGACTAATA AGACTAATTA
232381 ATTTAAAGTT TTTAAATCAC AACTGCGTGC AAAATAAATG GAAGTGGCAT GCTCGCCAAG
232441 TGTGCATGAG TGGTGTGCAT GGGAGACAGC ACGAAGCTAA TCCCACTCAT CTTGCAGGTT
232501 GCTCCATTTT TCTCTAAAA TCAGTAAGAC AGAAGCTGGT CAGATTATCA AGAGCCCTAG
232561 TTAAACACAG CAGTAGCATT TGGAAGGGGT TGCTCTCATT AGGCAGTGCC TGACCACAAC
232621 AAGAGATGAA CAAGCCCTGT ATCTGAAGCC ATCATGCCTA GTTATGGTCC CCGACTGTTC
232681 ATGATGCCTG GAAGGGAGGC CCCCTGCACC CTAGAAAGCT GGGTGGGTTT TACTGTCTGC
232741 TTTACTGCTA AAAACCCTCT TCTTTGGATC TGGACTTTAC CTCTATCTGA TTTTCTTTTC
232801 TAATATATGA TTTGGCACTG AGTCTGTAC TGCTGCTAAC TCAGCAAGTT TAGGGTCATT
232861 GCCCCATTGC CTCACAGAAA GAATTCATA GCTTCCAGCA TCCTCTCTCC TTCATTATAC
232921 TTTGATTTCA GCATTGCTAT TTTTCTCTT GGGTGTGCA GCTCTCTCTC TCCTTCCCAT
232981 GTCTTGTTGG TTTCTGCTA ACTCCTGCTT TTTTCTTTT TTTTCTTTT AGACGGAGTC
233041 TCGTTCTGTC ACCCAGGCTG GAGTGACAGT GCACAACTC GGCTCACTGC AACCTCCGCC
233101 TCCCGGGTTC AAGCTATTCT CCTGCCTCAG CCTCCCAAGT AGCTGGGACT ACAGGCCTC
233161 ACCACTATGC CCCACTAATT TTTGTATTTT TAGTATTGCT GTCATCAATC CACATGTCCA
233221 GAAGCACCTA GAACTCTAA TTCTTTGTAG GTATCAAACC CTAGGACTCT TTCCTCTAAT

Figure 9 (Page 72 of 74)

161/162

233281 CACAATATAT AATCCCTGAT TCCCAAACAC GGTCTTTTCA TATACATTTT CCACTGTACA
233341 TACTTTCTGA CCTGGAAAGC TCTTACACAA ACACGCCCTC CCCTAGGAAG CCTTTATAAA
233401 TGTTCCCAGG AAGAATCAGT CACCCAACAG TGTCTTGTC ACATCTTAGG TTCTACACCT
233461 TTATTTGTTT TATCTGAATG TAATCTCCA GAGGGTGTTA TCATCTTTT TTTTGAGATG
233521 GAGTCTTGCT TTGCTGCCCC GGCTGGAGTG CAGTGGCATG ATCTCGGCTC ACAGCAACCT
233581 CCACCTCCTG GGTTCAGTG ATTCTCCTGC CTCAGCCTCC TGAGTAGCTG GGATTACAGA
233641 CGTGTGTCAC CACACCTGGC TAATTTTGT ATTTTATAGTA GAGACAGGGT TTCACCGTGT
233701 TGGCAAGGCT TTCCTCGAAC TCCCAAACCT AGGTGATCCA CCCACCTCAG CCTCCCAAAG
233761 TGCTGGGATT ACAGGTGTGA GCCACCATGT CCAGCCCCAT CTTTCTCTT TAGTTTAGTT
233821 CTTAACAAAT AGTCTGACAC AAAGTGGATA TAACAATATT TTGAATTATG AATAACTAAA
233881 TGAATATTTT CAGATTTCTT GGTGCTCTCA AAGTTTATG TTACAAAAGA AAAACAAGTC
233941 TAAAATACCT GCCTCAAGTT TTTATCTGTA CTATGATTTT AAACCAAATA AAAAACAGGT
234001 GGGGTAAAAA CTGAAACAGG AAATACATAT AACTGAAAAA TTTTGGTATG TTAGTATGAT
234061 AATACTAGGT CATTTTCTCT GTTTCCTCAA CTTCATTTT TATAGCAATA AAAAGAAACA
234121 AGTAAATGTA TGTTAATTTA ATTTAAAAGA AGTAGTCTAC CATCTCTTCT GTTAAAAGA
234181 AAAAAGTATT TTAATAAATT ATCTCTGGAA GGATACACAG GGAACATTGC TCTGGTTTCT
234241 TCCAAGAGAG AAATGAGGAA CTAGAGAGCA TGGCCAAGTG GGGTTTGTCT TTTGTTTTTG
234301 TTTGTCTATC TGTTAGCTTT TTATTATTTT CTTTGTAGG TTTGAATTTT AAACACATA
234361 AATCTGTTAC ATGCTCATAA TAATAAGTTT AAAATAAAC TTTTGGCTGG GTGCAATGAC
234421 TTACACCTGT AATCCCAGCG CTTTGGGAAG CAGAGGTGGG AGGATACTTG AGGCCAGGAA
234481 TTTGAGATCA GCCTGGGCAA CATAGTGAGA CCCTGCCTCT GTAGAAATAA ACAAAAATTA
234541 GCTGGATATG GTGGTGCATG CTGTACTCC TAGCTACTTG GGAGGTTGAG GCAGGAGGAT
234601 CCTTTGAGTC CAGGAGTTTG AGGCTGCAGT GAGCTATAAT CACCCACTGC ACTATAGCAT
234661 GGGCAATAAG GTGAGAATT GTCTCAAAA AAAAAGGGGG GGGGGAACA AATAAATAAA
234721 TATAAACAAA ACTTTTGTCT CAAAATATGT AATATTTAGC ACTAAAGAAT TCTGAATTGT
234781 AGAGCTAAAA AGTTAATAAC TATTGTCTCC TTTAAAGAA TTGTTATCAA
234841 AGTATAATTT TTATCCAGAA AATCAGCAT ATCAGCAAGC TAACTTTCT CAAAATGACA
234901 TATCCATGTA ATTAGCTCCC AGGTAATTAG ACCCTTCATG GTCTCTTCTA GACCTCAGCT
234961 TAATCTAAAA ATTGGAAATT CAAAATGCTC CAAAATCTGC AACTTTTTGA ATGCTAACAT
235021 GATTCTCAAA GGAGTGCTCA TGGAGTATTT CAGATTTTGG ATTTTGGAT TTGAGATACT
235081 CAGTATAATG CAAACATTCC AAATCTGAAA AAATCTGAAA TACTTCTGGT TCTAAGCATA
235141 AGGGATACTC AACGTGTGTT AGCTAATTAG ACCCTTCATG GTCTCTTCTA GACCTCAGCT
235201 TCTTCAAGGT AACCTCTATC CTCACTCTA ATAGCATGAA CTTTCTGTT TTAGAATAAT
235261 TTGGATTTT AGGAAAGTTG CAAAGATAGT ACAAGACAG TACAGGAGAG TTCCCATATA
235321 TCTTTACCT AGCTTTCCCC CATTTGTTAG ATTTTACATT ATTATGATAC ATTTGTCAA
235381 TATAAGCAAC TCACATTGAT ACATGAACT CTATTAACCA AACCTTAGAC TTTATGTGGA
235441 TTTCACTACT GTTCCACTA ATGTTTCTT TCTGTTCCAA GGTCCAATG GGAATACCAC
235501 ACTGCATTTT CTTGTCATAT CTCCCTAGTC TTTTGTGTC TGTGACAATG TCTCAGTCTT
235561 TTCTTGCTTT TCATGACCTT AACAGTCTG AAGATCATTT GCTTTTTTT CATAATTACA
235621 CCGGAGTTAT AGATTTTTTG AAATAATACC ACAAGGGCAA AGGGCCCTTC TTGTCACATC
235681 ATTTTAGGGA GAACATGATA TCCACATGAC ATCACTGATA TTAACCTTCA TCATGTGGTT
235741 TAGGTAATGT TTCAGGTTTC TCTACTGCAA AGTGATTTT TTCCCTTAAT TTAGCCACC
235801 TGAACCTTAT AATTTGTTT TCTTCCATGA CTAATACTTT TGTATTATA GCTAAACTT
235861 CATTTGGGCC AAATCTTAGA TCATGTAAT TTTCTTCTAT ATTTATTCT AAAAGCTTGT
235921 AATGTTTGAT ACATTCTAAA AGATGTAATG TTTGATACAT TACATCTAGT CCTTTGATTT
235981 ATTTTATAGT ACTTTTGAT AAGGTGTGAG AGATGTCTCC AGTTTCACTT TATTAACACA
236041 TTGTGGTGTT CCAGTACTAT TTGTGCTAA GACTATCTTT TTCCATTGA TTACCTTTGC
236101 CTTAGTTGGC AATATTTTTG TTGGTTTATT TCTAGACTGT TTATCTCATT CCACTGATTT
236161 GTGTCTATCT TTTTGACAAA ACTGTTGATT ACAGTAAGCT TTGAAATAGT TCATTTTTTG
236221 TGTCAACTTG ACTGAGTCAG GGGATAACCA GCTATCTGGT TAAACATTAT TTCTGGCTGT
236281 GTTTGTGAGC GTGTTCTGAG ATGAGATTAG CCTTTGAATA GGTGATCCTA GTAAAGTAAA
236341 CTGTCTTTCC CAGTGTGGAT GGCATTATG CACCTGATAT TCAGGGTCTG AATAGAAGAA
236401 AAGGCAGAGG AAGGGGGAAT TTGGGCTTT TTTCTGCCT CACTGCTTGA GCTGGGACAT
236461 CTCATCTGGT CTCCTGCTCT TGAAGTGGGA TTTACATCAT CAGTTCCTCT GGTCTCAGG

Figure 9 (Page 73 of 74)

162/162

236521	CCTTCAGATT	CAGACTGAAT	CATACCACCA	GCTTTCCTGG	GTCTCCAGCT	TGCAGATTAC
236581	AGATCATGGG	ACTCCTCATC	TTCCATAAAT	GCATGAGCCA	ATTCAGTCTA	TGTCCTTGAA
236641	AACTGCCCCA	CTGCAGATTA	AGGCTTTTTT	CCACTAGGTG	AAATAAAGAA	GCTTGTTAGA
236701	CAGATTTCCC	TTCATCCAGT	GCCCTCTCCT	CTTTAAGTTA	CAACACATTG	GCTACACCTA
236761	AGTGCAGGGG	TGGGGATGAG	GGTATAGTCC	TCTTGTTTGC	TGAGAAGAGA	ACTGTATTGG
236821	GAAAGCTCTA	GAAGTGTTTG	ATACATACAT	AAACAAGGCA	TGGTTTTTGC	ACTTAATTTT
236881	ACATTACATT	TTTCCAGAA	AAAAAGGAAT	GTATAGGCAT	CACGTAAGT	TACTAGCTGG
236941	AGTCATTCTT	CCTGATTATC	AAAGGTAAAC	AGTTATTAAT	CCTATACCAA	GATGTCAAGG
237001	AGAAGTACTT	TTGGAACACA	AGGAATTCTC	TGGGAGTCCT	TACTACTCTC	AAGCCCAGTG
237061	AAAAAGTTAA	TGAAAAACTA	TAGTACCTTC	CTATAAGCTG	GATGACTAAT	TACCAGGCTC
237121	ATTTAGGAAT	TTGCCTTACC	AAGTAAAACA	TAAGGGCAGC	TGAGGTGCTG	ACTGAAGACA
237181	AATGGAGCAT	AGAATAAGAG	TAGTAAAGAA	TGCCAAAAAT	GCTGTCATGT	ATCCATTGAC
237241	AAAAGGAGCT	ATAAAGCCTT	TAGGTATTTT	CACACTTGCT	CTGTTACGTA	AATGTATGTG
237301	TGTGTGTGTG	TGTGTGTGTG	TGTGTG			

Figure 9 (Page 74 of 74)

SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

 International application No.
PCT/US97/17658

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C07H 21/04; C12Q 1/68; C12N 15/63, 15/85; C12P 21/02

US CL : 536/23.5; 435/6, 70.1, 325, 320.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5; 435/6, 70.1, 325, 320.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG'S BIOTECH cluster.

hemochromatosis, BTF1, BTF2, BTF3, BTF4, NTP-3, NTP-4, RoRet, butyrophilin, type I sodium transport

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	RUDDY, D.A. et al. A 1.1-Mb transcript map of the hereditary hemochromatosis locus. Genome Research. May 1997, Vol. 7, No. 5, pages 441-456, see entire document.	1-20, 22-77
X	FISCHER, L. et al. Cloning of the 62-kilodalton component of basic transcription factor BTF2. Science. 04 September 1992, Vol. 257, pages 1392-1395, see entire document.	28-33, 71
X	MARGOTTIN, F. et al. Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. Science. 25 January 1991, Vol. 251, pages 424-426, see entire document.	22-27, 70

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
I document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
I document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

20 JANUARY 1998

Date of mailing of the international search report

12 FEB 1998

 Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

F. PIERRE VANDERVEGT

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17658

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHENG, X.M. et al. Sequencing and expression of complementary DNA for the general transcription factor BTF3. Nature. 05 April 1990, Vol. 344, pages 556-559, see entire document.	34-39, 72
X	PANTEGHINI, M. Electrophoretic fractionation of 5'-nucleotidase. Clinical Chemistry. February 1994, Vol. 40, No. 2, pages 190-196, see entire document.	52-57, 75
X ---- A	BURT, M. J. et al. A 4.5-megabase YAC Contig and physical map over the hemochromatosis gene region. Genomics. 15 April 1996, Vol. 33, No. 2, pages 153-158, see entire document.	1-6 ---- 7-20, 22-77
A	VERNET, C. et al. Evolutionary study of multigenic families mapping close to the human MHC Class I region. J. Mol. Evol. November 1993, Vol. 37, No. 6, pages 600-612, see abstract in particular.	1-20, 22-77

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17658

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17658

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-20, drawn to polynucleotide sequences containing at least one polymorphic site, polypeptides encoded thereby, antibodies to said polypeptides and a method to determine the presence of the HFE gene mutation.

Group II, claim 21, drawn to the lymphoblastoid line atcc crl-12371.

Group III, claim(s) 22-27 and 70, drawn to BTF1 nucleic acids, gene products, vectors and antibodies.

Group IV, claim(s) 28-33 and 71, drawn to BTF2 nucleic acids, gene products, vectors and antibodies.

Group V, claim(s) 34-39 and 72, drawn to BTF3 nucleic acids, gene products, vectors and antibodies.

Group VI, claim(s) 40-45 and 73, drawn to BTF4 nucleic acids, gene products, vectors and antibodies.

Group VII, claim(s) 46-51 and 74, drawn to BTF5 nucleic acids, gene products, vectors and antibodies.

Group VIII, claim(s) 52-57 and 75, drawn to NPT3 nucleic acids, gene products, vectors and antibodies.

Group IX, claim(s) 58-63 and 76, drawn to NPT4 nucleic acids, gene products, vectors and antibodies.

Group X, claim(s) 64-69 and 77, drawn to RoRet nucleic acids, gene products, vectors and antibodies.

The inventions listed as Groups I-X do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I and III-X are drawn to physically different genes and their gene products and each therefore constitutes a separate invention. The lymphoblastoid cell line of Group II is not dependent upon the vectors of any of the Groups I and III-X and therefore constitutes a separate invention. Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68, 1/70, C07H 21/04, C12P 19/34		A1	(11) International Publication Number: WO 97/29212
			(43) International Publication Date: 14 August 1997 (14.08.97)
(21) International Application Number: PCT/US97/02102		(72) Inventors; and	
(22) International Filing Date: 7 February 1997 (07.02.97)		(75) Inventors/Applicants (for US only): GINGERAS, Thomas, A. [US/US]; 1528 Juniper Hill Drive, Encinitas, CA 92024 (US). MACK, David [US/US]; 2076 Monterey Avenue, Menlo Park, CA 94025 (US). CHEE, Mark, S. [AU/US]; 3199 Waverly Street, Palo Alto, CA 94306 (US). BERN, Anthony, J. [CA/US]; 570 South 12th Street, San Jose, CA 95112 (US). STRYER, Lubert [US/US]; 843 Sonoma Terrace, Stanford, CA 94035 (US). GHANDOUR, Ghassan [US/US]; 73 Palmer Lane, Atherton, CA 94027 (US). WANG, Ching [US/US]; 843 Cape Kennedy Drive, San Jose, CA 95133 (US).	
(30) Priority Data:			
60/011,339	8 February 1996 (08.02.96)	US	
60/012,631	1 March 1996 (01.03.96)	US	
08/629,031	8 April 1996 (08.04.96)	US	
60/017,765	15 May 1996 (15.05.96)	US	
(60) Parent Applications or Grants		(74) Agents: FITTS, Renee, A. et al.; Townsend and Townsend and Crew L.L.P., 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).	
(63) Related by Continuation			
US	60/017,765 (CIP)		
Filed on	15 May 1996 (15.05.96)		
US	08/629,031 (CIP)		
Filed on	8 April 1996 (08.04.96)		
US	60/012,631 (CIP)		
Filed on	1 March 1996 (01.03.96)		
US	60/011,339 (CIP)		
Filed on	8 February 1996 (08.02.96)		
(71) Applicant (for all designated States except US): AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
		Published With international search report.	
(54) Title: CHIP-BASED SPECIATION AND PHENOTYPIC CHARACTERIZATION OF MICROORGANISMS			
(57) Abstract			
<p>This invention provides oligonucleotide-based arrays and methods for speciating and phenotyping organisms, for example, using oligonucleotide sequences based on the <i>Mycobacterium tuberculosis rpoB</i> gene. The groups or species to which an organism belongs may be determined by comparing hybridization patterns of target nucleic acid from the organism to hybridization patterns in a database.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

CHIP-BASED SPECIATION AND PHENOTYPIC CHARACTERIZATION OF MICROORGANISMS

This application is a continuation-in-part of and claims the benefit of the priority dates of USSN 60/011,339, filed 08 Feb. 1996; 60/012,631, filed 01 March 1996; 08/629,031, filed 08 April 1996; and 60/017,765, filed 15 May 1996, the disclosures of which are specifically incorporated by reference in their entirety for all purposes.

BACKGROUND OF THE INVENTION

Copyright Notice

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the xerographic reproduction by anyone of the patent document or the patent disclosure in exactly the form it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

Field of the Invention

This invention relates to the identification and characterization of microorganisms.

Background of the Invention

Multidrug resistance and human immunodeficiency virus (HIV-1) infections are factors which have had a profound impact on the tuberculosis problem. An increase in the frequency of *Mycobacterium tuberculosis* strains resistant to one or more anti-mycobacterial agents has been reported, Block, et al., (1994) JAMA 271:665-671. Immunocompromised HIV-1 infected patients not infected with *M. tuberculosis* are frequently infected with *M. avium* complex (MAC) or *M. avium-M. intracellulare* (MAI) complex. These mycobacteria species are often resistant to the drugs used to treat *M. tuberculosis*. These factors have re-emphasized the importance for the accurate determination of drug sensitivities and mycobacteria species identification.

In HIV-1 infected patients, the correct diagnosis of the mycobacterial disease is essential since treatment of *M.*

tuberculosis infections differs from that called for by other mycobacteria infections, Hoffner, S.E. (1994) Eur. J. Clin. Microbiol. Inf. Dis. 13:937-941. Non-tuberculosis mycobacteria commonly associated with HIV-1 infections include *M. kansasii*, *M. xenopi*, *M. fortuitum*, *M. avium* and *M. intracellular*, Wolinsky, E., (1992) Clin. Infect. Dis. 15:1-12, Shafer, R.W. and Sierra, M.F. 1992 Clin. Infect. Dis. 15:161-162. Additionally, 13% of new cases (HIV-1 infected and non-infected) of *M. tuberculosis* are resistant to one of the primary anti-tuberculosis drugs (isoniazid [INH], rifampin [RIF], streptomycin [STR], ethambutol [EMB] and pyrazinamide [PZA] and 3.2% are resistant to both RIF and INH, Block, et al., JAMA 271:665-671, (1994). Consequently, mycobacterial species identification and the determination of drug resistance have become central concerns during the diagnosis of mycobacterial diseases.

Methods used to detect, and to identify *Mycobacterium* species vary considerably. For detection of *Mycobacterium tuberculosis*, microscopic examination of acid-fast stained smears and cultures are still the methods of choice in most microbiological clinical laboratories. However, culture of clinical samples is hampered by the slow growth of mycobacteria. A mean time of four weeks is required before sufficient growth is obtained to enable detection and possible identification. Recently, two more rapid methods for culture have been developed involving a radiometric, Stager, C.E. et al., (1991) J. Clin. Microbiol. 29:154-157, and a biphasic (broth/agar) system Sewell, et al., (1993) J. Clin. Microbiol. 29:2689-2472. Once grown, cultured mycobacteria can be analyzed by lipid composition, the use of species specific antibodies, species specific DNA or RNA probes and PCR-based sequence analysis of 16S rRNA gene (Schirm, et al. (1995) J. Clin. Microbiol. 33:3221-3224; Kox, et al. (1995) J. Clin. Microbiol. 33:3225-3233) and IS6110 specific repetitive sequence analysis (For a review see, e.g., Small et al., P.M. and van Embden, J.D.A. (1994) Am. Society for Microbiology, pp. 569-582). The analysis of 16S rRNA sequences (RNA and DNA) has been the most informative molecular approach to

identify *Mycobacteria* species (Jonas, et al., J. Clin. Microbiol. 31:2410-2416 (1993)). However, to obtain drug sensitivity information for the same isolate, additional protocols (culture) or alternative gene analysis is necessary.

To determine drug sensitivity information, culture methods are still the protocols of choice. *Mycobacteria* are judged to be resistant to particular drugs by use of either the standard proportional plate method or minimal inhibitory concentration (MIC) method. However, given the inherent lengthy times required by culture methods, approaches to determine drug sensitivity based on molecular genetics have been recently developed.

Table 1 lists the *M. tuberculosis* genes with which when mutated have been shown to confer drug resistance (other genes are known, e.g., the *pncA* gene). Of the drugs listed in Table 1, RIF and INH form the backbone of tuberculosis treatment. Detection of RIF resistance in *M. tuberculosis* is important not only because of its clinical and epidemiological implications but also because it is a marker for the highly threatening multidrug resistant phenotype (Telenti, et al. (1993) The Lancet 341:647-650). Of the drug resistances listed in Table 1, decreased sensitivity to RIF is the best understood on a genetic basis.

Table 1

M. tuberculosis Genes with Mutations Which Confer Drug Resistance

Drug	Gene	Size (bp)	Gene Product
RIF	<i>rpoB</i>	3,534	β -subunit of RNA polymerase
INH	<i>katG</i>	2,205	catalase-peroxidase
INH-ETH	<i>inhA</i>	810	fatty acid biosynthesis
STR	<i>rpsL</i>	372	ribosomal protein S12
	<i>rrs</i>	1,464	16S rRNA
FQ	<i>gyrA</i>	2,517	DNA gyrase A subunit

Because resistance to RIF in *E. coli* strains was observed to arise as a result of mutations in the *rpoB* gene,

Telenti, et al., id., identified a 69 base pair (bp) region of the *M. tuberculosis* *rpoB* gene as the locus where RIF resistant mutations were focused. Kapur, et al., (1995) Arch. Pathol. Lab. Med. 119:131-138, identified additional novel mutations in the *M. tuberculosis* *rpoB* gene which extended this core region to 81 bp. In a detailed review on antimicrobial agent resistance in mycobacteria, Musser (Clin. Microbiol. Rev., 8:496-514 (1995)), summarized all the characterized mutations and their relative frequency of occurrence in this 81 bp region of *rpoB*. Missense mutations comprise 88% of all known mutations while insertions (3 or 6 bp) and deletions (3, 6 and 9 bp) account for 4% and 8% of the remaining mutations, respectively. Approximately 90% of all RIF resistant tuberculosis isolates have been shown to have mutations in this 81 bp region. The remaining 10% are thought possibly to involve genes other than *rpoB*.

For the above reasons, it would be desirable to have simpler methods which identify and characterize microorganisms, such as *Mycobacteria*, both at the phenotypic and genotypic level. This invention fulfills that and related needs.

SUMMARY OF THE INVENTION

The present invention provides systems, methods, and devices for characterizing and identifying organisms. In one aspect of the invention, a method for identifying a genotype of a first organism, comprising:

(a) providing an array of oligonucleotides at known locations on a substrate, said array comprising probes complementary to reference DNA or RNA sequences from a second organism;

(b) hybridizing a target nucleic acid sequence from the first organism to the array; and

(c) based on an overall hybridization pattern of the target to the array, identifying the genotype of the first organism, and optionally identifying a phenotype of the first organism.

Another aspect of the invention provides a method for identifying the genotype and/or phenotype of an organism by comparing a target nucleic acid sequence from a first organism coding for a gene (or its complement) to a reference sequence coding for the same gene (or its complement) from a second organism, the method comprising:

(a) hybridizing a sample comprising the target nucleic acid or a subsequence thereof to an array of oligonucleotide probes immobilized on a solid support, the array comprising:

a first probe set comprising a plurality of probes, each probe comprising a segment of nucleotides exactly complementary to a subsequence of the reference sequence, the segment including at least one interrogation position complementary to a corresponding nucleotide in the reference sequence;

(b) determining which probes in the first probe set bind to the target nucleic acid or subsequence thereof relative to their binding to the reference sequence, such relative binding indicating whether a nucleotide in the target sequence is the same or different from the corresponding nucleotide in the reference sequence;

(c) based on differences between the nucleotides of the target sequence and the reference sequence identifying the phenotype of the first organism;

(d) deriving one or more sets of differences between the reference sequence and the first organism; and

(e) comparing the set of differences to a data base comprising sets of differences correlated with speciation of organisms to identify the genotype of the first organism.

Another aspect of the invention provides a method for identifying the genotype and/or phenotype of an organism by comparing a target nucleic acid sequence from a first organism coding for a gene (or its complement) to a reference sequence coding for the same gene (or its complement) from a second organism, the method comprising:

(a) hybridizing a sample comprising the target nucleic acid or a subsequence thereof to an array of

oligonucleotide probes immobilized on a solid support, the array comprising:

a first probe set comprising a plurality of probes, each probe comprising a segment of nucleotides exactly complementary to a subsequence of the reference sequence, the segment including at least one interrogation position complementary to a corresponding nucleotide in the reference sequence, wherein each interrogation position corresponds to a nucleotide position in the reference or target sequence;

(b) determining a hybridization intensity from each probe;

(c) plotting the hybridization intensities versus the nucleotide position corresponding to the probe from which the hybridization intensity was determined to derive a target plot of hybridization intensity;

(d) repeating steps (a) - (c) with the target sequence replaced by the reference sequence, to derive a baseline plot of the reference sequence; and

(e) comparing the target plot to the baseline plot to identify the genotype and/or phenotype of the organism.

Another aspect of the invention provides an array of oligonucleotide probes immobilized on a solid support, the array comprising:

a first probe set comprising a plurality of probes, each probe comprising a segment of nucleotides exactly complementary to a subsequence of a reference sequence, the segment including at least one interrogation position complementary to a corresponding nucleotide in the reference sequence;

wherein the reference sequence is a gene from *Mycobacterium tuberculosis*.

Another aspect of the invention provides a method of identifying the presence of a nucleic acid polymorphism in a patient sample, comprising the steps of:

(a) determining the difference between the hybridization intensities of a nucleic acid sequence from the patient sample and a corresponding nucleic acid sequence from

a wild type sample to an array of reference nucleic acid probes;

(b) deriving ratios of the difference in (a) to the hybridization intensity of the wild type sample for each base position corresponding to each reference nucleic acid probe; and

(c) identifying the presence of a polymorphism at a base position corresponding to a reference probe if the ratio in (b) for the base position corresponding to the reference probe is greater than or equal to an assigned value.

Another aspect of the invention provides a computer program product that identifies the presence of a nucleic acid polymorphism in a patient sample, comprising:

computer code that determines the difference between the hybridization intensities of a nucleic acid sequence from the patient sample and a corresponding nucleic acid sequence from a wild type sample to an array of reference nucleic acid probes;

computer code that derives ratios of the difference to the hybridization intensity of the wild type sample for each base position corresponding to each reference nucleic acid probe;

computer code that identifies the presence of a polymorphism at a base position corresponding to a reference probe if the ratio for the base position corresponding to the reference probe is greater than or equal to an assigned value; and

a computer readable medium that stores the computer codes.

Another aspect of the invention provides, in a computer system, a method of assigning an organism to a group, comprising the steps of:

inputting groups of a plurality of known nucleic acid sequences, the plurality of known nucleic acid sequences being from known organisms;

inputting hybridization patterns for the plurality of known nucleic acid sequences, each hybridization pattern indicating hybridization of subsequences of the known nucleic

acid sequence to subsequences of a reference nucleic acid sequence;

inputting a hybridization pattern for a sample nucleic acid sequence from the organism indicating hybridization of subsequences of the sample nucleic acid sequence to subsequences of the reference nucleic acid sequence;

comparing the hybridization pattern for the sample nucleic acid sequence to the hybridization patterns for the plurality of known nucleic acid sequences; and

assigning a particular group to which the organism belongs according to the group of at least one of the known nucleic acid sequences that has a hybridization pattern that most closely matches the hybridization pattern of the sample nucleic acid sequence at specific locations.

Another aspect of the invention provides a computer program product that assigns an organism to a group, comprising:

computer code that receives as input groups of a plurality of known nucleic acid sequences, the plurality of known nucleic acid sequences being from known organisms;

computer code that receives as input hybridization patterns for the plurality of known nucleic acid sequences, each hybridization pattern indicating hybridization of subsequences of the known nucleic acid sequence to subsequences of a reference nucleic acid sequence;

computer code that receives as input a hybridization pattern for a sample nucleic acid sequence from the organism indicating hybridization of subsequences of the sample nucleic acid sequence to subsequences of the reference nucleic acid sequence;

computer code that compares the hybridization pattern for the sample nucleic acid sequence to the hybridization patterns for the plurality of known nucleic acid sequences;

computer code that assigns a particular group to which the organism belongs according to the groups of at least one of the known nucleic acid sequences that has a

hybridization pattern that most closely matches the hybridization pattern of the sample nucleic acid sequence at specific locations; and

a computer readable medium that stores the computer codes.

Another aspect of the invention provides, in a computer system, a method of assigning groups to which organisms belong utilizing a generic probe array, comprising the steps of:

inputting hybridization intensities for a plurality of isolates, the hybridization intensities indicating hybridization affinity between the isolate and the generic probe array;

selecting hybridization intensities that have the most variance across the plurality of isolates; and

assigning each of the plurality of isolates to a group according to the selected hybridization intensities.

Another aspect of the invention provides a computer program product that assigns groups to which organisms belong utilizing a generic probe array, comprising the steps of:

computer code that receives as input hybridization intensities for a plurality of isolates, the hybridization intensities indicating hybridization affinity between the isolate and the generic probe array;

computer code that selects hybridization intensities that have the most variance across the plurality of isolates;

computer code that assigns a group to each of the plurality of isolates according to the selected hybridization intensities; and

a computer readable medium that stores the computer codes.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1: Basic tiling strategy. The figure illustrates the relationship between an interrogation position (I) and a corresponding nucleotide (n) in the reference sequence, and between a probe from the first probe set and corresponding probes from second, third and fourth probe sets.

Fig. 2: Segment of complementarity in a probe from the first probe set.

Fig. 3: Incremental succession of probes in a basic tiling strategy. The figure shows four probe sets, each having three probes. Note that each probe differs from its predecessor in the same set by the acquisition of a 5' nucleotide and the loss of a 3' nucleotide, as well as in the nucleotide occupying the interrogation position.

Fig. 4: Exemplary arrangement of lanes on a chip. The chip shows four probe sets, each having five probes and each having a total of five interrogation positions (I1-I5), one per probe.

Fig. 5: Strategies for detecting deletion and insertion mutations. Bases in brackets may or may not be present.

Fig. 6: Shows the light directed synthesis of oligonucleotide probes on a substrate.

Fig. 7: Shows the synthesis of a combinatorial array all possible tetranucleotide oligomers on a chip.

Fig. 8: A schematic diagram of target preparation.

Fig. 9: A tiling strategy for sequence determination.

Fig. 10: A mismatch profile for an octamer based chip.

Fig. 11: A hypothetical six-class tree based classification system. The numbers underneath the terminal nodes are the class assignments as determined by this classifier.

Fig. 12: An image of the Mtb *rpoB* chip analysis of the 700 bp region of the *rpoB* gene from an *M. tuberculosis* isolate.

Fig. 13: Shared single nucleotide polymorphisms of seven *Mycobacterium* species.

Fig. 14: Unique (species-specific) single nucleotide polymorphisms of seven *Mycobacterium* species.

Fig. 15A and 15B: Hybridization patterns and bar code fingerprint representations of seven *Mycobacterium* species.

Fig. 16: Bar code fingerprint representations of seven *M. gordonae* clinical isolates and the core fingerprint derived therefrom.

Fig. 17: Plot of hybridization intensity vs. nucleotide position using *M. gordonae* as target on an *Mycobacterium tuberculosis* *rpoB* chip. The bottom panel shows the sequences of the *rpoB* genes of *M. tuberculosis* and *M. gordonae* with the position of difference outlined in black.

Fig. 18: Plot of hybridization intensity vs. nucleotide position using other *Mycobacterium* species as target on an *Mycobacterium tuberculosis* *rpoB* chip.

Fig. 19: Plots of hybridization intensity vs. nucleotide position using *Mycobacterium* species as target on an *Mycobacterium tuberculosis* *rpoB* chip overlaid on the corresponding plot for *Mycobacterium tuberculosis*.

Fig. 20A-20D and 21: Plots of hybridization intensity vs. nucleotide position of an unknown patient sample compared to plots of known *Mycobacterium* species as target on an *Mycobacterium tuberculosis* *rpoB* chip.

Fig. 22: Plots of hybridization intensity vs. nucleotide position of *Mycobacterium gordonae* isolates as target on an *Mycobacterium tuberculosis* *rpoB* chip compared to a reference ATCC isolate.

Fig. 23 (A, B): Design of a tiled array.

Fig. 24: Effect and positional dependence of a single base mismatch on hybrid stability using the MT1 DNA chip. The sequences of the perfect match probe and each A:A single base mismatch probe are shown. The results of five independent experiments are plotted.

Fig. 25 (A, B, C): Detection of base differences in a 2.5 kb region of human mitochondrial DNA between a sample and reference target by comparison of scaled p^0 hybridization intensity patterns.

Fig. 26 (A, B): Detection of deletion sequences of human mitochondrial DNA.

Fig. 27 (A, B, C): Hybridization of 16.3 kb of a mitochondrial target to chip with the entire mitochondrial genome.

Fig. 28 (A, B): (A) Overlay of hybridization intensities of Exon 12 of the MSH2 gene from a patient sample and from a wild type sample. (B) Plot of hybridization intensity differences greater than 0.25 between patient sample and wild type sample as a function of base position.

Fig. 29: Plot of hybridization intensity differences greater than 0.25 between patient sample and wild type sample as a function of base position for Exon 13 and Exon 16 of the MLH1 gene.

Fig. 30: Plot of hybridization intensity differences greater than 0.25 between patient sample and wild type sample as a function of base position for Exon 12 of the MSH2 gene.

Fig. 31: Plot of hybridization intensity differences greater than 0.25 between patient sample and wild type sample as a function of base position for Exon 5 of the p53 gene.

Fig. 32: Computer that may be utilized to execute software embodiments of the present invention.

Fig. 33: A system block diagram of a typical computer system that may be used to execute software embodiments of the invention.

Fig. 34: A high level flowchart of identifying the presence of a polymorphism in a nucleic acid sequence from a patient sample.

Fig. 35: A high level flowchart of a method of identifying a species within a genus to which an organism belongs.

Fig. 36: A high level flowchart of a method of identifying species within a genus to which organisms belong.

Fig. 37: A hierarchical clustering of isolates of *Mycobacterium*.

DESCRIPTION OF THE PREFERRED EMBODIMENT

This invention provides methods, compositions and devices for identifying the group or species of an organism and obtaining functional phenotypic information about the

organism based on genotypic analysis of one or more genomic regions of the organism. In one embodiment, the method compares a target nucleic acid sequence from the organism coding for a gene (or its complement) to a reference sequence coding for the same gene (or its complement).

In principle, a reference sequence from any genomic region of the organism can be used. When phenotypes are being identified, it will be understood by one of skill in the art that mutations within that region will affect the phenotypic trait which is being characterized. Genotyping, by contrast, only requires that a polymorphism, which may or may not code for a mutation, be present. The reference sequence can be from a highly polymorphic region, a region of intermediate polymorphic complexity or in some cases, a highly conserved region. Highly polymorphic regions are typically more informative when doing speciation analysis. The method disclosed herein is readily applicable to using reference sequences from highly polymorphic regions, though in certain cases one may prefer to use a reference sequence from a highly conserved region within the organism, since this reduces the mathematical complexity of the deconvolution required of the overall hybridization patterns observed during the analysis. In this context, a "highly conserved region" of a organism refers to a degree of conservation at the genotypic level of greater than 50%, preferably greater than 75%, and more preferably greater than 90%. A particularly useful reference sequence is the 700 bp *rpoB* gene from *Mycobacterium tuberculosis* (Mt), since it is well defined. Other useful reference sequences include 16SrRNA, *rpoB* gene, *katG* gene, *inhA* gene, *gyrA* gene, 23SrRNA gene, *rrs* gene, *pncA* gene, and *rpsL* gene. Furthermore, an 81 bp segment within this gene contains all the known mutations which code for rifampacin resistance in *M. tuberculosis*. The invention is particularly useful for phenotypic and genotypic characterization of microorganisms. In this context, the term "microorganism" refers to bacteria, fungi, protozoa or viruses.

The invention finds particular utility in assaying biological samples. The term "biological sample", as used

herein, refers to a sample obtained from an organism or from components (e.g., cells) of an organism. The sample may be of any biological tissue or fluid. Frequently the sample will be a "clinical sample" which is a sample derived from a patient.

5 Such samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. Biological samples may also include sections of tissues such as frozen sections taken for histological
10 purposes.

Frequently, it is desirable to amplify the nucleic acid sample prior to hybridization. Suitable amplification methods include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., *PCR Protocols. A guide to*
15 *Methods and Application*. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, *Genomics*, 4: 560 (1989), Landegren, et al., *Science*, 241: 1077 (1988) and Barringer, et al., *Gene*, 89: 117 (1990), transcription amplification (Kwoh, et al., *Proc. Natl. Acad.*
20 *Sci. USA*, 86: 1173 (1989)), and self-sustained sequence replication (Guatelli, et al., *Proc. Nat. Acad. Sci. USA*, 87: 1874 (1990)).

In a preferred embodiment, the hybridized nucleic acids are detected by detecting one or more labels attached to
25 the sample nucleic acids. The labels may be incorporated by any of a number of means well known to those of skill in the art. However, in a preferred embodiment, the label is simultaneously incorporated during the amplification step in the preparation of the sample nucleic acids. Thus, for
30 example, polymerase chain reaction (PCR) with labeled primers or labeled nucleotides will provide a labeled amplification product. In a preferred embodiment, transcription amplification, as described above, using a labeled nucleotide (e.g. fluorescein-labeled UTP and/or CTP) incorporates a label
35 into the transcribed nucleic acids.

Alternatively, a label may be added directly to the original nucleic acid sample (e.g., mRNA, polyA mRNA, cDNA, etc.) or to the amplification product after the amplification

is completed. Means of attaching labels to nucleic acids are well known to those of skill in the art and include, for example nick translation or end-labeling (e.g. with a labeled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore).

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., DynabeadsTM), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

An oligonucleotide probe array complementary to the reference sequence or subsequence thereof is immobilized on a solid support using one of the display strategies described below. For the purposes of clarity, much of the following description of the invention will use probe arrays derived from the *Mycobacterium rpoB* gene as an example; however it should be recognized, as described previously, that probe arrays derived from other genes may also be used, depending on the phenotypic trait being monitored, the availability of suitable primers and the like.

Initially, target nucleic acids derived from *Mycobacterium* species having *rpoB* genes of known sequence and known drug resistance mutations are screened against a solid phase probe array derived from sequences complementary to the *Mycobacterium tuberculosis rpoB* gene (the Mtb *rpoB* chip). The known sequences are either available from the literature or can be independently established by another method, such as

dideoxynucleotide sequencing. The overall hybridization pattern observed with each these species is compared to the overall hybridization pattern observed with *Mycobacterium tuberculosis* and differences between the two hybridization patterns are derived. A sample derived from the *Mycobacterium tuberculosis* (Mt) used as the reference sequence, being exactly complementary to the probe set(s) on the solid support, will bind to all the probes. Samples derived from other organisms, which contain one or more polymorphisms at the genotypic level, will not display similar binding. The observed patterns will vary as a function of the variation in the sequences of the *rpoB* genes of the individual species. Subsequences identical to Mt will generate hybridization subpatterns identical to the subpattern observed with Mt for that corresponding subsequence. Subsequences which differ from Mt will generate hybridization subpatterns which differ from the Mt subpattern for that corresponding subsequence. Thus, the overall hybridization pattern observed with a particular species allows one to identify regions of the *rpoB* gene of that species which differ from that of Mt.

The presence of a different hybridization pattern in a specified region of the substrate can be correlated with a probability that the target nucleic acid is from a specific species. In the idealized case, the differential hybridization pattern in a single region will allow species identification. This can occur when one or more polymorphisms in that region are uniquely associated with a specific species. More frequently however, such an unique one-to-one correspondence is not present. Instead, differential hybridization patterns (i.e., relative to the reference sequence) are observed in multiple regions, none of which will bear an unique correspondence to a particular species. However, each differential hybridization pattern will be associated with a probability of the organism being screened belonging to a particular species (or not) or carrying a particular phenotypic trait (or not). As a result, detection of an increasing number of these sets of differences allows one to classify the organism with an increasing level of

confidence. Other algorithms can be used to derive such composite probabilities from the detection of multiple sets of differences. Therefore, the overall hybridization pattern, which is the aggregate of all the differential hybridizations observed at all regions of the substrate, allows one to assign with high confidence, the speciation and/or phenotype of the organism.

When a single probe set is being used on the substrate, it will usually not be able to define the differences in sequence between the target and the reference sequence, absent additional knowledge about the target. Multiple probe sets can be used, as with the tiling strategies disclosed herein and described in more detail in PCT publication WO95/11995. In some cases, the differences will be definable, i.e., the different nucleotide responsible for the different hybridization pattern will be known. In other cases the difference will not be definable, i.e., all one will know is that a polymorphism is present in that region. However, this is primarily a function of the probe array used on the chip. If necessary the sample can be screened against a different probe array to assign the polymorphism present in that region. Since the point mutations which confer antibiotic resistance, for example to rifampacin, for Mt are frequently known, the presence of a change in the hybridization pattern in the region where the point mutation occurs signals the presence of a rifampacin resistant species. It will be apparent that this technique is not limited to identifying drug resistance. Any phenotypic trait whose variation has been mapped to mutations in a particular genomic region can be identified by this method. Representative and nonlimiting examples include the presence of toxin and pathogenic markers.

It is important to recognize that this method provides more than the ability to identify genotypic variations and thus phenotype by hybridization. Analysis of hybridization patterns of a single genomic region of a microorganism with the Mt rpoB chip also provides, as

explained below, a method of identifying the species of the microorganism.

This chip based screening method allows one to build up a data base of hybridization patterns corresponding to different species. Some regions of the hybridization pattern will be shared among subsets of the species because their sequences in regions corresponding to those hybridizations are identical. Other regions of the hybridization pattern will differ between two species because the sequence corresponding to those hybridizations are different. In all cases, the sequences of the *rpoB* gene of the unknown species are being compared to the corresponding sequence of Mt. Differences in the hybridization pattern of a particular species to the pattern observed with Mt as sample, can be correlated to the presence of a polymorphism at a particular point in the sequence of that species. Some polymorphisms will be definable, i.e., one will know not only that the nucleotide at that position differs from that of Mt, but one will also know the identity of that nucleotide. Some polymorphisms will be unique to the particular species, i.e., species-specific; they will be present in that particular species and not in any other species. Other polymorphisms will be shared, i.e., they will not be unique to a particular species. Certain subsets of species will have the polymorphism and others will not. However, each of these polymorphisms can be assigned to its particular subset of species. Therefore, the presence of a shared polymorphism, despite not indicating with certainty that the sample being screened contains a particular species, increases the probability that one species of that particular subset of species is present.

The hybridization pattern of a particular sample can be represented as a "bar code" in which the individual lines of the bar code indicate the presence of a polymorphism relative to Mt. This invention provides a method of screening large numbers of individual species and thus deriving information on the polymorphisms present in those species. Each individual line can be assigned a probability of being associated with different species. In this fashion, a data

base can be built up in which increasing numbers of polymorphisms can be associated with the different species. As one will recognize, the presence of an unique species-specific polymorphism will allow the immediate identification of a sample as being a particular species. However, even the presence of shared polymorphisms among several species will allow species identification. In the simplest case, each species can be assigned a "fingerprint" of shared polymorphisms, i.e., that species and isolates of that species will possess a particular collection of shared polymorphisms. However, it is not necessary for one to be able to assign an unique "fingerprint" pattern of shared polymorphisms to a species in order to be able to identify that species. As long as one can correlate the presence of a particular polymorphism or subset of polymorphisms with a probability of the sample being a particular species, the detection of increasing numbers of such polymorphisms allows one to predict with increasing probability the speciation of the sample, i.e., as one observes more and more such polymorphisms associated with a particular species, the confidence level of the sample being that particular species increases. Standard mathematical algorithms can be used to make this prediction. Therefore, once the data base is sufficiently large, the lack of an "unique" fingerprint for a species becomes irrelevant. Typically, the mathematical algorithm will make a call of the identity of the species and assign a confidence level to that call. One can determine the confidence level (>90%, >95% etc.) that one desires and the algorithm will analyze the hybridization pattern and either provide an identification or not. Occasionally, the call may be that the sample may be one of two, three or more species, in which case a specific identification will not be possible. However, one of the strengths of this technique is that the rapid screening made possible by the chip-based hybridization allows one to continuously expand the data base of patterns and polymorphisms to ultimately enable the identification of species previously unidentifiable due to lack of sufficient information.

Analysis of an increasing number of isolates of a known species will allow one to build up a fingerprint that is characteristic of that species. However, it is important to note that as the total number of analyzed isolates for each species is increased, it is unlikely that a single and unique core fingerprint will define a *Mycobacterium* species. Rather, it is expected that any particular isolate of a *Mycobacterium* species will have a subset of all possible fingerprints. Identification of the *Mycobacterium* species based on a fingerprint pattern will require a classification analysis, such as by using a tree-based classification algorithm as described below, built upon a collected database consisting of species specific and shared single nucleotide polymorphisms (SNPs) and fingerprints. Thus, the chip-based method of determining hybridization patterns disclosed herein allows one to both build up a data base of polymorphisms associated with a particular species and use that data base to identify the speciation and phenotypic characteristic of an unknown sample from a single hybridization experiment.

It should also be recognized that since this technique rests on differences in hybridization patterns, this method of speciation does not rest on knowing the actual identity of the polymorphism. The hybridization pattern relative to Mt will differ as long as a nucleotide at a particular point in the sequence differs from that of Mt. The exact nature of the substitution, insertion or deletion, e.g., A to T, C or G is less important than the fact that the nucleotide is not A (assuming for the purposes of illustration that Mt carries an A at that position). It is not necessary that the sample be sequenced in order to identify its speciation.

A second layer of confidence can be added to the initial determination by analyzing whether the differences in hybridization patterns are shared or unique. If the species identified is supposed to have either a shared or unique polymorphism at a particular site and the chip has in fact detected such a polymorphism, then one can be more confident in the initial determination.

Both identification and phenotyping can be accomplished based on genotypic determinations of a single region of the mycobacteria genome in place of analysis of two genomic regions (the *rpoB* and the 16S rRNA genes). Two generic implications can be derived from the successful demonstration of the use of high density oligonucleotide arrays for mycobacteria identification and antibiotic drug sensitivity. First, other genes affecting drug sensitivity can be encoded on the high density oligonucleotide arrays (see Table 1) and hybridization patterns for each of these additional genes can be used to confirm and provide confidence measurements for fingerprints derived from the *rpoB* gene. Second, the same chip-based strategy could be employed for other eubacteria species which simultaneously could provide genotypic information concerning important clinical phenotypes (e.g., toxin and pathogen marker genes) as well as identification information.

The preceding discussion has used the Mt *rpoB* gene as an example. It should be recognized that this method is generally applicable to other microorganisms and reference sequences derived from other genomic regions, such as, for example, the human mitochondrial DNA sequence and the MT DNA sequence. The length of a reference sequence can vary widely from a full-length genome, to an individual chromosome, episome, gene, component of a gene, such as an exon or regulatory sequences, to a few nucleotides. A reference sequence of between about 2, 5, 10, 20, 50, 100, 5000, 1000, 5,000 or 10,000, 20,000 or 100,000 nucleotides is common. Sometimes only particular regions of a sequence are of interest. In such situations, the particular regions can be considered as separate reference sequences or can be considered as components of a single reference sequence, i.e., the microbial genome.

A reference sequence can be any naturally occurring, mutant, consensus sequence of nucleotides, RNA or DNA. For example, sequences can be obtained from computer data bases, publications or can be determined or conceived *de novo*. Usually, a reference sequence is selected to show a high

degree of sequence identity to envisaged target sequences. Often, particularly, where a significant degree of divergence is anticipated between target sequences, more than one reference sequence is selected. Combinations of wild-type and mutant reference sequences are employed in several applications of the tiling strategy.

Fig. 32 illustrates an example of a computer system that may be used to execute software embodiments of the present invention. Fig. 32 shows a computer system 100 which includes a monitor 102, screen 104, cabinet 106, keyboard 108, and mouse 110. Mouse 110 may have one or more buttons such as mouse buttons 112. Cabinet 106 houses a CD-ROM drive 114, a system memory and a hard drive (see Fig. 33) which may be utilized to store and retrieve software programs incorporating code that implements the present invention, data for use with the present invention, and the like. Although a CD-ROM 116 is shown as an exemplary computer readable storage medium, other computer readable storage media including floppy disks, tape, flash memory, system memory, and hard drives may be utilized. Cabinet 106 also houses familiar computer components such as a central processor, system memory, hard disk, and the like.

Fig. 33 shows a system block diagram of computer system 100 that may be used to execute software embodiments of the present invention. As in Fig. 32, computer system 100 includes monitor 102 and keyboard 108. Computer system 100 further includes subsystems such as a central processor 102, system memory 120, I/O controller 122, display adapter 124, removable disk 126 (e.g., CD-ROM drive), fixed disk 128 (e.g., hard drive), network interface 130, and speaker 132. Other computer systems suitable for use with the present invention may include additional or fewer subsystems. For example, another computer system could include more than one processor 102 (i.e., a multi-processor system) or a cache memory.

Arrows such as 134 represent the system bus architecture of computer system 100. However, these arrows are illustrative of any interconnection scheme serving to link the subsystems. For example, a local bus could be utilized to connect the central processor to the system memory and display

adapter. Computer system 100 shown in Fig. 33 is but an example of a computer system suitable for use with the present invention. Other configurations of subsystems suitable for use with the present invention will be readily apparent to one
5 of ordinary skill in the art.

The methods of this invention employ oligonucleotide arrays which comprise probes exhibiting complementarity to one or more selected reference sequences whose sequence is known. Typically, these arrays are immobilized in a high density
10 array ("DNA on chip") on a solid surface as described in U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070, WO 92/10092 and WO 95/11995, each of which is incorporated herein by reference.

Various strategies are available to order and
15 display the oligonucleotide probe arrays on the chip and thereby maximize the hybridization pattern and sequence information derivable regarding the target nucleic acid. Exemplary display and ordering strategies are described in PCT patent publication No. WO 94/12305, incorporated herein by
20 reference. For the purposes of fuller description, a brief description of the basic strategy is described below.

The basic tiling strategy provides an array of immobilized probes for analysis of target sequences showing a high degree of sequence identity to one or more selected
25 reference sequences. The strategy is illustrated for an array that is subdivided into four probe sets, although it will be apparent that satisfactory results are obtained from one probe set (i.e., a probe set complementary to the reference sequence as described earlier).

30 A first probe set comprises a plurality of probes exhibiting perfect complementarity with a selected reference sequence. The perfect complementarity usually exists throughout the length of the probe. However, probes having a segment or segments of perfect complementarity that is/are
35 flanked by leading or trailing sequences lacking complementarity to the reference sequence can also be used. Within a segment of complementarity, each probe in the first probe set has at least one interrogation position that

corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference sequence, when the probe and reference sequence are aligned to maximize complementarity between the two. If a probe has more than one interrogation position, each corresponds with a respective nucleotide in the reference sequence. The identity of an interrogation position and corresponding nucleotide in a particular probe in the first probe set cannot be determined simply by inspection of the probe in the first set. As will become apparent, an interrogation position and corresponding nucleotide is defined by the comparative structures of probes in the first probe set and corresponding probes from additional probe sets.

In principle, a probe could have an interrogation position at each position in the segment complementary to the reference sequence. Sometimes, interrogation positions provide more accurate data when located away from the ends of a segment of complementarity. Thus, typically a probe having a segment of complementarity of length x does not contain more than $x-2$ interrogation positions. Since probes are typically 9-21 nucleotides, and usually all of a probe is complementary, a probe typically has 1-19 interrogation positions. Often the probes contain a single interrogation position, at or near the center of probe.

For each probe in the first set, there are, for purposes of the present illustration, up to three corresponding probes from three additional probe sets. Fig. 1 illustrates the basic "tiling" strategy of the invention. Thus, there are four probes corresponding to each nucleotide of interest in the reference sequence. Each of the four corresponding probes has an interrogation position aligned with that nucleotide of interest. Usually, the probes from the three additional probe sets are identical to the corresponding probe from the first probe set with one exception. The exception is that at least one (and often only one) interrogation position, which occurs in the same position in each of the four corresponding probes from the four probe

sets, is occupied by a different nucleotide in the four probe sets. For example, for an A nucleotide in the reference sequence, the corresponding probe from the first probe set has its interrogation position occupied by a T, and the
5 corresponding probes from the additional three probe sets have their respective interrogation positions occupied by A, C, or G, a different nucleotide in each probe. Of course, if a probe from the first probe set comprises trailing or flanking sequences lacking complementarity to the reference sequences
10 (see Fig. 2), these sequences need not be present in corresponding probes from the three additional sets. Likewise corresponding probes from the three additional sets can contain leading or trailing sequences outside the segment of complementarity that are not present in the corresponding
15 probe from the first probe set. Occasionally, the probes from the additional three probe set are identical (with the exception of interrogation position(s)) to a contiguous subsequence of the full complementary segment of the
20 corresponding probe from the first probe set. In this case, the subsequence includes the interrogation position and usually differs from the full-length probe only in the omission of one or both terminal nucleotides from the termini of a segment of complementarity. That is, if a probe from the first probe set has a segment of complementarity of length n ,
25 corresponding probes from the other sets will usually include a subsequence of the segment of at least length $n-2$. Thus, the subsequence is usually at least 3, 4, 7, 9, 15, 21, or 25 nucleotides long, most typically, in the range of 9-21 nucleotides. The subsequence should be sufficiently long or
30 hybridization conditions such to allow a probe to hybridize detectably more strongly to a variant of the reference sequence mutated at the interrogation position than to the reference sequence.

The probes can be oligodeoxyribonucleotides or
35 oligoribonucleotides, or any modified forms of these polymers that are capable of hybridizing with a target nucleic sequence by complementary base-pairing. Complementary base pairing means sequence-specific base pairing which includes e.g.,

Watson-Crick base pairing as well as other forms of base pairing such as Hoogsteen base pairing. Modified forms include 2'-O-methyl oligoribonucleotides and so-called PNAs, in which oligodeoxyribonucleotides are linked via peptide bonds rather than phosphodiester bonds. The probes can be attached by any linkage to a support (e.g., 3', 5' or via the base). 3' attachment is more usual as this orientation is compatible with the preferred chemistry for solid phase synthesis of oligonucleotides.

The number of probes in the first probe set (and as a consequence the number of probes in additional probe sets) depends on the length of the reference sequence, the number of nucleotides of interest in the reference sequence and the number of interrogation positions per probe. In general, each nucleotide of interest in the reference sequence requires the same interrogation position in the four sets of probes.

Consider, as an example, a reference sequence of 100 nucleotides, 50 of which are of interest, and probes each having a single interrogation position. In this situation, the first probe set requires fifty probes, each having one interrogation position corresponding to a nucleotide of interest in the reference sequence. The second, third and fourth probe sets each have a corresponding probe for each probe in the first probe set, and so each also contains a total of fifty probes. The identity of each nucleotide of interest in the reference sequence is determined by comparing the relative hybridization signals at four probes having interrogation positions corresponding to that nucleotide from the four probe sets.

In some reference sequences, every nucleotide is of interest. In other reference sequences, only certain portions in which variants (e.g., mutations or polymorphisms) are concentrated are of interest. In other reference sequences, only particular mutations or polymorphisms and immediately adjacent nucleotides are of interest. Usually, the first probe set has interrogation positions selected to correspond to at least a nucleotide (e.g., representing a point mutation) and one immediately adjacent nucleotide. Usually, the probes

in the first set have interrogation positions corresponding to at least 3, 10, 50, 100, 1000, 20,000, 100,000, 1,000,000, 10,000,000, or more contiguous nucleotides. The probes usually have interrogation positions corresponding to at least 5, 10, 30, 50, 75, 90, 99 or sometimes 100% of the nucleotides in a reference sequence. Frequently, the probes in the first probe set completely span the reference sequence and overlap with one another relative to the reference sequence. For example, in one common arrangement each probe in the first probe set differs from another probe in that set by the omission of a 3' base complementary to the reference sequence and the acquisition of a 5' base complementary to the reference sequence. Figure 3 illustrates an incremental succession of probes in a basic tiling strategy.

The number of probes on the chip can be quite large (e.g., 10^5 - 10^6). However, often only a relatively small proportion (i.e., less than about 50%, 25%, 10%, 5% or 1%) of the total number of probes of a given length are selected to pursue a particular tiling strategy. For example, a complete set of octomer probes comprises 65,536 probes; thus, an array of the invention typically has fewer than 32,768 octomer probes. A complete array of decamer probes comprises 1,048,576 probes; thus, an array of the invention typically has fewer than about 500,000 decamer probes. Often arrays have a lower limit of 25, 50 or 100 probes and as many probes as 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , etc. probes. The arrays can have other components besides the probes such as linkers attaching the probes to a support.

Some advantages of the use of only a proportion of all possible probes of a given length include: (i) each position in the array is highly informative, whether or not hybridization occurs; (ii) nonspecific hybridization is minimized; (iii) it is straightforward to correlate hybridization differences with sequence differences, particularly with reference to the hybridization pattern of a known standard; and (iv) the ability to address each probe independently during synthesis, using high resolution photolithography, allows the array to be designed and

optimized for any sequence. For example the length of any probe can be varied independently of the others.

For conceptual simplicity, the probes in a set are usually arranged in order of the sequence in a lane across the chip, although this arrangement is not required. For example, the probes can be randomly distributed on the chip. A lane contains a series of overlapping probes, which represent or tile across, the selected reference sequence (see Fig. 3). The components of the four sets of probes are usually laid down in four parallel lanes, collectively constituting a row in the horizontal direction and a series of 4-member columns in the vertical direction. Corresponding probes from the four probe sets (i.e., complementary to the same subsequence of the reference sequence) occupy a column. Each probe in a lane usually differs from its predecessor in the lane by the omission of a base at one end and the inclusion of additional base at the other end as shown in Fig. 3. However, this orderly progression of probes can be interrupted by the inclusion of control probes or omission of probes in certain columns of the array. Such columns serve as controls to orient the chip, or gauge the background, which can include target sequence nonspecifically bound to the chip.

The probes sets are usually laid down in lanes such that all probes having an interrogation position occupied by an A nucleotide form an A-lane, all probes having an interrogation position occupied by a C nucleotide form a C-lane, all probes having an interrogation position occupied by a G nucleotide form a G-lane, and all probes having an interrogation position occupied by a T (or U) form a T lane (or a U lane). Note that in this arrangement there is not a unique correspondence between probe sets and lanes. Thus, the probe from the first probe set is laid down in the A-lane, C-lane, A-lane, A-lane and T-lane for the five columns in Fig. 4. The interrogation position on a column of probes corresponds to the position in the target sequence whose identity is determined from analysis of hybridization to the probes in that column. Thus, I_1 - I_5 , respectively correspond to N_1 - N_5 in Fig. 4. The interrogation position can be anywhere in

a probe but is usually at or near the central position of the probe to maximize differential hybridization signals between a perfect match and a single-base mismatch. For example, for an 11 mer probe, the central position is the sixth nucleotide.

5 Although the array of probes is usually laid down in rows and columns as described above, such a physical arrangement of probes on the chip is not essential. Provided that the spatial location of each probe in an array is known, the data from the probes can be collected and processed to
10 yield the sequence of a target irrespective of the physical arrangement of the probes on a chip. In processing the data, the hybridization signals from the respective probes can be reassorted into any conceptual array desired for subsequent data reduction whatever the physical arrangement of probes on
15 the chip.

 A range of lengths of probes can be employed in the chips. As noted above, a probe may consist exclusively of a complementary segments, or may have one or more complementary segments juxtaposed by flanking, trailing and/or intervening
20 segments. In the latter situation, the total length of complementary segment(s) is more important than the length of the probe. In functional terms, the complementary segment(s) of the first probe sets should be sufficiently long to allow the probe to hybridize detectably more strongly to a reference
25 sequence compared with a variant of the reference including a single base mutation at the nucleotide corresponding to the interrogation position of the probe. Similarly, the complementary segment(s) in corresponding probes from additional probe sets should be sufficiently long to allow a
30 probe to hybridize detectably more strongly to a variant of the reference sequence having a single nucleotide substitution at the interrogation position relative to the reference sequence. A probe usually has a single complementary segment having a length of at least 3 nucleotides, and more usually at
35 least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 30 bases exhibiting perfect complementarity (other than possibly at the interrogation

position(s) depending on the probe set) to the reference sequence.

In some chips, all probes are the same length. Other chips employ different groups of probe sets, in which case the probes are of the same size within a group, but
5 differ between different groups. For example, some chips have one group comprising four sets of probes as described above in which all the probes are 11 mers, together with a second group comprising four sets of probes in which all of the probes are
10 13 mers. Of course, additional groups of probes can be added. Thus, some chips contain, e.g., four groups of probes having sizes of 11 mers, 13 mers, 15 mers and 17 mers. Other chips have different size probes within the same group of four probes. In these chips, the probes in the first set can vary
15 in length independently of each other. Probes in the other sets are usually the same length as the probe occupying the same column from the first set. However, occasionally different lengths of probes can be included at the same column position in the four lanes. The different length probes are
20 included to equalize hybridization signals from probes depending on the hybridization stability of the oligonucleotide probe at the pH, temperature, and ionic conditions of the reaction.

The length of a probe can be important in
25 distinguishing between a perfectly matched probe and probes showing a single-base mismatch with the target sequence. The discrimination is usually greater for short probes. Shorter probes are usually also less susceptible to formation of secondary structures. However, the absolute amount of target
30 sequence bound, and hence the signal, is greater for larger probes. The probe length representing the optimum compromise between these competing considerations may vary depending on *inter alia* the GC content of a particular region of the target DNA sequence, secondary structure, synthesis efficiency and
35 cross-hybridization. In some regions of the target, depending on hybridization conditions, short probes (e.g., 11 mers) may provide information that is inaccessible from longer probes (e.g., 19 mers) and vice versa. Maximum sequence information

can be read by including several groups of different sized probes on the chip as noted above. However, for many regions of the target sequence, such a strategy provides redundant information in that the same sequence is read multiple times from the different groups of probes. Equivalent information can be obtained from a single group of different sized probes in which the sizes are selected to maximize readable sequence at particular regions of the target sequence.

Some chips provide an additional probe set specifically designed for analyzing deletion mutations. The additional probe set comprises a probe corresponding to each probe in the first probe set as described above. However, a probe from the additional probe set differs from the corresponding probe in the first probe set in that the nucleotide occupying the interrogation position is deleted in the probe from the additional probe set, as shown in Figure 5. Optionally, the probe from the additional probe set bears an additional nucleotide at one of its termini relative to the corresponding probe from the first probe set (shown in brackets in Fig. 5). The probe from the additional probe set will hybridize more strongly than the corresponding probe from the first probe set to a target sequence having a single base deletion at the nucleotide corresponding to the interrogation position. Additional probe sets are provided in which not only the interrogation position, but also an adjacent nucleotide is deleted.

Similarly, other chips provide additional probe sets for analyzing insertions. For example, one additional probe set has a probe corresponding to each probe in the first probe set as described above. However, the probe in the additional probe set has an extra T nucleotide inserted adjacent to the interrogation position. See Fig. 5 (the extra T is shown in a square box). Optionally, the probe has one fewer nucleotide at one of its termini relative to the corresponding probe from the first probe set (shown in brackets). The probe from the additional probe set hybridizes more strongly than the corresponding probe from the first probe set to a target sequence having an A insertion to the left of nucleotide "n"

of the reference sequence in Fig. 5. Similar additional probe sets can be constructed having C, G or A nucleotides inserted adjacent to the interrogation position.

Usually, four such additional probe sets, one for each nucleotide, are used in combination. Comparison of the hybridization signal of the probes from the additional probe sets with the corresponding probe from the first probe set indicates whether the target sequence contains an insertion. For example, if a probe from one of the additional probe sets shows a higher hybridization signal than a corresponding probe from the first probe set, it is deduced that the target sequence contains an insertion adjacent to the corresponding nucleotide (n) in the target sequence. The inserted base in the target is the complement of the inserted base in the probe from the additional probe set showing the highest hybridization signal. If the corresponding probe from the first probe set shows a higher hybridization signal than the corresponding probes from the additional probe sets, then the target sequence does not contain an insertion to the left of corresponding position ("n" in Fig. 5) in the target sequence.

Other chips provide additional probes (multiple-mutation probes) for analyzing target sequences having multiple closely spaced mutations. A multiple-mutation probe is usually identical to a corresponding probe from the first set as described above, except in the base occupying the interrogation position, and except at one or more additional positions, corresponding to nucleotides in which substitution may occur in the reference sequence. The one or more additional positions in the multiple mutation probe are occupied by nucleotides complementary to the nucleotides occupying corresponding positions in the reference sequence when the possible substitutions have occurred.

Another aspect of the invention derives hybridization patterns from a chip with a first probe set comprising a plurality of probes of perfect complementarity to the reference sequence, and optionally, one or more additional probe sets, each additional set comprising probes

corresponding to a probe in the first set with an interrogation position for a nucleotide of interest. The probes in the additional probe sets differ from their corresponding probes in the first probe set by having a different nucleotide in the interrogation position. The overall hybridization is derived by plotting the maximum hybridization intensity observed from target hybridization to the group of probes consisting of a probe in the first probe set and its corresponding probes in the additional probe sets versus the nucleotide position of the target being interrogated by this group of probes. Thus, in this method, the probes are grouped according to groups in which all the probes in a particular group interrogate a common nucleotide position in the sequence being analyzed. These groups are referred to as groups of interrogatory probes. For example, with reference to Fig. 4, the first column of probes with interrogation position I^1 is interrogating position n^1 , the second column of probes with interrogation position I^2 is interrogating position n^2 and so on. In the case described above in Fig. 1, where the corresponding probe from the first probe set has a T nucleotide in the interrogation position and the corresponding probes from the other three probe sets have a C, G, and A nucleotide respectively, there would be a total of four probes interrogating the particular nucleotide of interest at that position of the sample sequence being analyzed. One measures the highest of the intensities observed from each of these probes and assigns that measured value as being the maximum hybridization intensity corresponding to that position of the sample sequence. This determination is then repeated iteratively for the remaining nucleotide positions of the sample sequence being analyzed, allowing one to obtain a plot of hybridization intensity vs. nucleotide position.

It should be recognized that it is not necessary that there be additional sets of corresponding probes which interrogate all four possible nucleotide polymorphisms at a particular position. The method described immediately above measures both the maximum hybridization intensity at a

particular position and also how that maximum intensity changes from that position to the next adjacent position as one scans or "tiles through" the sequence of the target. Therefore, the chip can use a single probe set complementary to the reference sequence; multiple probe sets each of which
5 to the reference sequence; multiple probe sets each of which interrogate a particular position in the target by varying the corresponding nucleotide of interest at that position; or even additional probe sets which are of different lengths to the first and additional probe sets comprising the first set of
10 groups of interrogatory probes.

For example, one can use a chip with a single probe set complementary to the reference that tiles across the reference sequence. In this case, despite there being only one probe which interrogates a particular nucleotide position
15 of the sample, one plots the hybridization intensity of that particular probe as a function of the nucleotide position being interrogated. In this case, there is no "maximum" hybridization intensity because each position is being "interrogated" by just one probe. However, one can still
20 derive from the image plots of hybridization intensity as a function of nucleotide position of the sequence being analyzed and build up databases which correlate genotype with the derived plot. This method, using one set of probes (complementary to the reference sequence) is described in U.S.
25 patent application Serial No. 08/531,137, filed October 16, 1995. A plot obtained from such a method is shown in Fig. 17. This entire plot is derived from the image gathered from a single hybridization experiment.

As the sample sequence being analyzed varies, the
30 shape of this plot will also vary. For the purposes of clarity of explanation the following discussion uses a chip with probes complementary to a reference sequence from the *Mycobacterium tuberculosis rpoB* gene and the plot of maximum hybridization intensity derived from the image observed when
35 the reference sequence (i.e., from an *M. tuberculosis* sample) is hybridized to the chip is called the baseline or reference plot (or pattern). Target sequences from species of *Mycobacterium* other than *tuberculosis* will give plots of

different shape. Hybridization experiments with targets of known speciation thus provides a database in which each of these differently shaped plots is correlated with speciation or other genotypic feature, which then in turns allows one to predict the presence of a phenotype. It should be apparent that any gene of interest can be tiled across the chip and that the hybridization pattern derived from the image on the chip from any other sample suspected of containing that gene or polymorphic variant thereof can be used to detect the presence or absence of a particular variant of that gene in the sample.

Fig. 17 shows the plot of hybridization intensity as a function of position being interrogated along the reference sequence for the case where the reference sequence is from the *Mycobacterium tuberculosis* *rpoB* gene (the Mtb chip described in the Examples) and the target is *M. gordonae*. Fig. 18 shows similar plots obtained with other *Mycobacterium* species. It is also noteworthy that species different to *M. tuberculosis* produce differences in hybridization intensity even from segments of the sequence which are identical to *M. tuberculosis*. Thus this method allows one to derive information even from subsequences that are identical to *M. tuberculosis*. As will be apparent, each species produces a characteristic pattern. One can pick the pattern obtained with the reference sequence, in this case, the Mt *rpoB*, as being the baseline (or reference) pattern and overlay the pattern from the target over the pattern from the reference to detect differences from the reference. Fig. 19 shows such an overlay of patterns from different targets versus Mtb as observed on an Mtb chip. As one expects, when the target is *M. tuberculosis* (bottom panel) the overlay is perfect, whereas when the target is not *M. tuberculosis* differences are present. Each of these patterns is thus a "fingerprint" for that particular species. Once a database of fingerprints is established for one can compare the corresponding pattern obtained from an unknown target to either conclusively identify the target as being a particular species, or exclude the possibility of that target being any one of the species

represented in the data base. Figs. 20A-20D show such a comparison of an unknown patient sample to fingerprints from four *Mycobacterium* species showing a match and thus identifying the unknown as being *M. gordonae*. Figure 21 shows a similar identification of two other samples (6 and 7, previously incorrectly identified as *M. avium* by another technique) as *M. xenopi* and *M. intracellulare*.

As the above discussion indicates, there are several ways of plotting hybridization intensity versus nucleotide position of the sample, all of which provide patterns which are characteristic of a genotypic difference. As such, this invention is not limited using plots derived by the specific protocols disclosed herein. As explained earlier, identification of genotype and genotypic differences also allow the prediction of a phenotype.

It should also be recognized that the sequence used to generate the baseline pattern against which the target pattern is compared need not be that derived from a sample in which the reference sequence was hybridized to the chip. Any other sample that is related to the target sample may be used since the method compares differences between the baseline pattern and the pattern from the unknown target.

High density oligonucleotide arrays may be utilized to detect drug resistance conferring mutations using information gathered from gene regions utilized to identify species of isolates within the genus *Mycobacterium*. For example, the 700 base pairs of the *rpoB* gene of *Mycobacterium* may be utilized to detect mutations that confer resistance to rifampin and to detect polymorphisms which allow for the identification of *Mycobacterium* species.

Table 1B indicates the total polymorphic variation observed among the nine non-tuberculosis species compared to *M. tuberculosis* within the 700 base pairs of *rpoB*. With any of these non-tuberculosis species there are both species specific (base positions where observed polymorphisms found only with that species) and share (base positions which have polymorphisms found in some of the isolates of that species and some isolates of other *Mycobacterium* species). Virtually

all of these polymorphisms have never been previously described and constitute useful and important markers for the identification of their corresponding species.

Table 1B

Mycobacteria Polymorphic Analysis

Species	Total Polymorphisms	Species Specific	Shared
M. avium	72	3	69
M. chelonae	106	8	48
M. fortuitum	103	21	81
M. gordonae	102	26	76
M. intracellulare	59	3	56
M. kansasii	84	12	72
M. scrofulaceum	62	2	60
M. smegmatis	101	10	91
M. xenopii	73	13	60

16SrRNA sequences are commonly utilized to identify species of *Mycobacterium*. However, analysis of the hybridization pattern using the *rpoB* gene indicates that there are some isolates that have been misclassified. For example, two *Mycobacterium* isolates received from the California Public Health department, 96-1761 and 95-1760, were indicated as *M. avium* isolates. When the *rpoB* gene is utilized, it was determined that the most similar match was with *M. intracellulare* (a close relative of *M. avium*).

The following will describe an embodiment that identifies the species within a genus to which an organism belongs. However, the process is generally applicable to assigning groups to organisms, where the groups may be species, subspecies, phenotypes, genotypes, and the like. Accordingly, the description that follows illustrates one embodiment of the invention.

Fig. 35 shows a computer-implemented flowchart of a method of identifying a species within a genus to which an organism belongs. At step 300, species of nucleic acid sequences from known organisms are input. These nucleic acid

sequences will be called "known nucleic acid sequences."
Additionally, at step 302, hybridization patterns for the
known nucleic acid sequences. The hybridization patterns
indicate the hybridization affinity of subsequences of the
5 known nucleic acid sequences to subsequences of a reference
nucleic acid sequence. For example, the subsequences of the
reference nucleic acid sequence may be portions of nucleic
acid probes on a chip.

At step 304, a database of the species and
10 hybridization patterns of the known nucleic acid sequences may
be generated. As with other steps, this step is optional but
may make identifying species more efficient.

The system compares the hybridization pattern of a
sample nucleic acid sequence to the hybridization patterns for
15 the known nucleic acid sequences at step 308, which may be
optionally stored in a database. At step 308, the system
determines the species of the organism from which the sample
nucleic acid was obtained according to the hybridization
pattern of the known nucleic acid sequences that most closely
20 matches the hybridization pattern of the sample nucleic acid
at specific locations. Although an overall pattern matching
technique may be utilized, one may also analyze species
specific polymorphic locations and/or shared polymorphic
locations. Additionally, it may be a combination of
25 hybridization patterns that are utilized to identify the
species of the sample nucleic acid sequence.

Comparing the hybridization patterns may be done in
any number of known techniques. In a preferred embodiment,
linear regression is utilized across all or selected base
30 positions to normalize the hybridization intensities. A
regression coefficient from the linear regression is then
utilized to measure the closeness of the hybridization
intensities of the hybridization patterns and therefore, the
nucleic acid sequences. Additionally, depending on how
35 closely the hybridization pattern of the sample nucleic acid
matches a hybridization pattern of the known nucleic acid
sequences, the system may calculate a probability that the

identified species for the sample is correct as indicated at step 310.

Again referring to Fig. 19, the figure shows plots of hybridization intensities of *Mycobacteria* species. A DNA assay was designed for *Mycobacteria tuberculosis* (Mtb) as the chip wild-type sequence. This chip will be referred to as the Mtb chip to indicate that the chip was tiled for Mtb. In other words, in addition to other probes, there are probes that are perfectly complementary to Mtb at sequential base locations. These probes will be referred to as the wild-type probes or probes complementary to the reference sequence.

Mtb was hybridized to the Mtb chip and the hybridization intensities of the wild-type probes (here measured as a logarithmic function of the photon counts) vs. the base position is shown in the bottom graph identified as "Mtb vs. *M. tuberculosis*." The graph illustrates an example of a hybridization pattern for Mtb. As indicated by the title of the graph, the graph actually shows the Mtb hybridization intensity pattern vs. itself so that there are actually two hybridization patterns superimposed on each other. In the following paragraphs, the Mtb sequence will be identified as the reference sequence (i.e., is typically a known sequence).

There are many species of *Mycobacteria*. Numerous species were hybridized to the Mtb chip and the graphs in Fig. 19 show the hybridization intensities of the wild-type probes (again measured as a logarithmic function of the photon counts) vs. the base position. In addition to the hybridization pattern for the *Mycobacteria* species, each graph also shows the hybridization pattern for the reference sequence, Mtb.

Although 80% of the bases of the different *Mycobacteria* species may be the same, each species generates a unique hybridization pattern or footprint. A sample sequence which is known to be a *Mycobacteria* species (e.g., from previous base calling algorithms or dideoxy sequencing) may be similarly hybridized to the Mtb chip. The hybridization pattern that results may be compared to the hybridization

patterns of known sequences to determine the identity of the sample sequence.

The *Mycobacteria* species themselves (or other species) may have enough similarities that the base calling algorithm is able to identify the sample as a *Mycobacteria* species. The species may also have enough differences that this method is able to identify the species according to the hybridization pattern.

Although in this example, the chip-wild type sequence and the reference sequence were the same sequence, different sequences may be utilized. The hybridization patterns discussed were generated by the hybridization intensities of the wild-type probes. However, hybridization patterns may be generated other ways including utilizing hybridization intensities of the highest intensity probe at each base position. Additionally, the method may be utilized on other species or even unrelated nucleic acid sequences to identify a sample sequence.

Typically, the hybridization differences observed between different species are large, whereas, as expected, the differences between different isolates of the same species are smaller. Therefore, one can set the cut off of the discriminating pattern matching function to whatever predetermined level is desired, depending on whether one is attempting to assign speciation or track an isolate. Figure 22 shows the patterns observed with different isolates of *M. gordonae* and their comparison to a single isolate of *M. gordonae* (ATCC isolate).

It should be noted that derivation of hybridization intensity vs. nucleotide position patterns and their correlation with patterns of known identity does not require that one identify the base present at particular position of the target or sequence the target. Instead, one determines the maximum hybridization intensity observed from any of the one or more probes which are interrogating for the presence of nucleotide identical to that of the reference sequence at the corresponding position in the target and plots how this changes as a function of base position. The pattern thus

obtained is compared to a database of patterns from organisms of known speciation to establish the presence or absence of match. Thus, there is no necessity to "call" or identify any of the bases in the target sequence in order to make an assignment.

Once differences between the target sequence pattern and the baseline reference pattern are established, these differences can be used in the same manner as the presence of differences in nucleotide sequence between target and reference to derive probabilities that the presence of a certain level of difference in hybridization intensity at a particular position indicate a certain species or genotype. All the observed differences from the reference can then be combined to give a composite probability of the sample being of a particular species or genotype. Thus, derivation of these patterns of hybridization also allows the use of the "bar code" type of identification method described earlier. As Fig. 17 shows, using patterns derived from hybridization intensity allows one to obtain information from the entire sequence, not just the regions where the sequence of the target differs from that of the reference.

One advantage of this method of pattern matching is that provided the same set of probes is used on a chip, one can use different chips at different times and with different concentrations of sample to make the assignment because each different species will produce the same and invariant hybridization pattern. For example, one does not need to derive a control pattern from the reference sequence simultaneous with the analysis of the target to comparing the two patterns (target vs. reference sequence), since the control pattern is invariant and the pattern matching looks at the relative changes in maximum hybridization intensities between succeeding base positions along the sequence. Thus, factors such as amplification conditions and sample concentration which would affect hybridization at all sites equally can be normalized during the analysis.

One will recognize that this method of using oligonucleotide arrays with such pattern matching techniques

is generally applicable to reference sequences other than the *rpoB* gene and can be used to detect any differences between a target and reference sequence from any gene. By way of example, and not limitation, the reference sequence can be from a gene coding for an HIV gene, breast cancer (BRCA-1) or for cystic fibrosis. Software is used to plot the hybridization intensities and compare the pattern so derived to a pattern from the reference or other sequence to establish differences between the target and reference, identity or lack thereof of target to sequence in the database of patterns.

Polynucleotide sequence can be represented as an assembly of overlapping oligonucleotides. Therefore, an array consisting of the set of complementary oligonucleotides to a specific sequence can be used to determine the identity of a target sequence, quantitate the amount of the target, or detect differences between the target sequence and a reference. Many different arrays can be designed for these purposes. One such design, termed a tiled array, is depicted schematically in Fig. 23A.

The use of a tiled array of probes to read a target sequence is illustrated in Fig. 23B. A $p^{15,7}$ (15-mers varied at position 7 from the 3' end) tiled array was designed and synthesized against MT1, a cloned sequence containing 1,311 bases spanning the D-loop, or control region of the human mitochondrial DNA. The upper image panel of Fig. 23B shows a portion of the fluorescence image of mt¹ fluorescein labelled RNA hybridized to the array. The base sequence can be read by comparing the intensities of the four probes within each column. For example, the column labelled 16,493 consists of the four probes, 3' TGACATAGGCTGTAG (SEQ ID NO:1), 3' TGACATCGGCTGTAG (SEQ ID NO:2), 3' TGACATGGGCTGTAG (SEQ ID NO:3), 3' TGACATTGGCTGTAG (SEQ ID NO:4). The probe with the strongest signal is the probe with the A substitution (A 301, C 57, G 135, T 110 counts), identifying the base at position 16,493 as a U (complementary to the A probe) in the RNA transcript. Continuing the process, the rest of the sequence can be read directly from the hybridization intensities.

The detection of a single base polymorphism is shown in the lower image panel of Fig. 23B. The target hybridized in the lower panel is MT2, which differs from MT1 in this region by a T to C transition at position 16,493.

5 Accordingly, the probe with the G substitution shows the strongest signal. Since the tiled array was designed to MT1, neighboring probes which overlap 16,493 are also affected by the change. Because 15-mer probes are used, a total of 15 columns, or 60 probes, are affected by a single base change in
10 the target. In the $p^{15,7}$ array, probes in the 8 positions to the left and 6 positions to the right of the probe set interrogating the mutation have an additional mismatch to the target. The result is a characteristic "footprint", or loss of signal in the probes flanking a mutation position,
15 reminiscent of the U shaped curve of Fig. 24. (The data shown in Fig. 24 are for 8 mer probes, but we have been able to discriminate single base end position mismatches from perfect matches even using 20 mers). Of the four interrogation probes at each position, signal loss is greatest from the probe
20 designed to have zero mismatches to MT1. We identify the set of these designed probes as $p^{015,7}$ or simply P^0 . In the other three probe sets, designated p^1 , the MT1 signal is already low as a result of the single base mismatch at the interrogation position.

25 Comparative Hybridization and Multi-Color Detection

Patterns of signal intensities and their differences resulting from mismatches, such as the example shown in Fig. 23B, can be used to advantage in sequence analysis. The loss
30 of hybridization signal from P^0 is a powerful indicator of sequence difference between reference and target. This information is best obtained by hybridizing both the reference and the target sequence simultaneously to the same array. In order to extract the maximum amount of useful information from
35 a simple tiled array, we developed a two-color labelling and detection scheme, allowing us to use as an internal standard the hybridization of a reference sample of known sequence (Fig. 25). The reference is labelled with phycoerythrin (red)

and the unknown target with fluorescein (green). This approach minimizes or eliminates experimental variability during the fragmentation, hybridization, washing, and detection steps. A further advantage is that the sample and reference targets are in competition, enhancing mismatch discrimination.

It is also possible to perform the experiment by hybridizing the reference and unknown to two different chips in parallel under identical conditions. In this case, only a single label is required. Using either approach, differences between two related sequences can be identified from a straightforward comparison of the scaled hybridization patterns of the p^0 probes. Differences in p^0 intensities resulting from a polymorphism extend over a number of positions and correlate with probe length and substitution position (Fig. 25). This characteristic large-scale pattern is more robust and easier to recognize than an intensity difference at a single position. Since the amplitude of each p^0 signal is sequence and mismatch dependent, the actual size and shape of footprints is variable. Thus, sequences can be identified by directly comparing hybridization amplitude signatures, rather than by comparing analyzed sequences, which may contain embedded errors of interpretation. Hybridization pattern analysis may provide advantages over other methods of detecting sequence variation, or in some cases may be useful in conjunction with them.

Polymorphism Screening of the Cytochrome b Gene and Control Region

We have shown how a p^0 probe set in conjunction with a reference hybridization pattern can be used to analyze sequences over much larger spans than 600 bp, and how single base polymorphisms can be read using a tiled array. We combined the two approaches and used array hybridization patterns to perform automated basecalling from complex target sequences. The combination was useful in overcoming most difficulties. For example, some sequences can cross-hybridize significantly, particularly when the target is long and

repetitive on a fine scale, or GC rich, even locally. By analyzing targets in terms of differences from a reference of known sequence, many potentially confusing signals could be disregarded because they were the same in both samples. A
5 second limitation is that if two or more polymorphisms occur within a single probe span, the resulting destabilization tends to reduce the accuracy of sequence interpretation, although the existence of a change can easily be inferred from the loss of signal (Fig. 25). We adopted an approach that
10 simply identifies such regions for further analysis, rather than attempting to read them directly.

After applying an automated basecalling algorithm, which uses all four interrogation probes for each position and compares the reference and sample hybridization intensities
15 the derived sequences were separated into two categories: one that could be read directly with high accuracy, and a second that required further analysis for definitive sequence assignment. The first category was defined as having a derived sequence with no more than a single mismatch with each
20 p^0 probe, and agreement between the derived sequence and p^0 footprint patterns.

The p^0 intensity footprints were detected in the following way: the reference and sample intensities were normalized, and R , the average of $\log_{10} (p^0 \text{ reference}/p^0 \text{ sample})$ over a window of 5 positions, centered at the base of
25 interest, calculated for each position in the sequence. To normalize the sample probe intensities to the reference intensities, a histogram of the base 10 logarithm of the intensity ratios for each pair of probes was constructed. The
30 histogram has a mesh size of 0.01, and was smoothed by replacing the value at each point with the average number of counts over a five-point window centered at that point. The highest value in the histogram was located, and the resulting intensity ratio was taken to be the most probable calibration
35 coefficient. Footprints were detected as regions having at least 5 contiguous positions with a reference or sample intensity of at least 50 counts above background, and an R value in the top 10th percentile for the experiment. At 205

polymorphic sites, where the sample was mismatched to p^0 , the mean R value was 1.01, with a standard deviation of 0.57. At 35,333 non polymorphic sites (i.e., where both reference and sample had a perfect match to p^0) the mean R value was -0.05, with a standard deviation of 0.25.

The second category had a derived sequence with multiple mismatches and/or disagreement between derived sequence and p^0 patterns. For example, the region of ief007 shown in Fig. 25A would fall into the first category, if the sequence were called correctly. A false positive basecall would lack a footprint or would result in the prediction of multiple mismatched probes, and be flagged in either case. False negatives are detected by the presence of a footprint despite a "wild-type" basecall.

An example of a false negative basecall resulting from the use of a limited probe set is shown in Fig. 26A. In this case, there is a $(CA)_n$ length polymorphism, where $n = 4$ in MT2 and $n = 5$ in the reference, MT1. The array is designed to read $(CA)_5$, but the MT2 target hybridizes sufficiently well to be read as "wild type". However, a footprint is detected, and therefore, the region is flagged for further analysis. Some differences in hybridization patterns are secondary effects of a change elsewhere in the target. An example that is likely due to a sequence-specific difference in target secondary structure is shown in Fig. 26A. This example shows that the interpretation of difference patterns is not always straightforward. In general, however, the difference patterns provide a substantial amount of additional information that aids sequence analysis.

We analyzed a 2.5 kb region of MT DNA spanning the $tRNA^{Glu\ Thr}$, cytochrome b, $tRNA^{Thr}$, $tRNA^{Pro}$, control region and $tRNA^{Phe}$ DNA sequences. These sequences have very different functions ranging from protein coding to tRNA structure to regulatory, and should therefore provide a good comparative basis for evaluating the different mutation rates of MT DNA. The $p^{20.9}$ tiling array was used to analyze a total of 12 samples containing 180 (0.59%) base substitutions relative to MT1. The results are presented in Table 1A. Of the 30,582 bp

analyzed, base substitutions were read in 98% of the sequence with > 99% accuracy without user intervention. No false positive calls were made. The remaining 2% of sequence was flagged for further analysis. This indicates a very high rate of accuracy, which was obtained for the analysis of 2.5 kb of sequence at one time. Thus, while the more mature gel-based technology was able to read clustered and length polymorphisms, hybridization to a 4N tiled array yielded comparable results over most of the sequence with considerably less effort.

Table 1A. Sequence analysis results.

Polymorphic Sites ^a					Non-polymorphic Sites			
	TOTAL		Called ^b		TOTAL		Called	
Mismatches ^c	0	≥1	0	≥1	0	≥	0	≥
All Positions	134	46	130	35	26883	3465	26883	3457
Unflagged Regions	126	1	126	0	26732	3020	26732	3020
Flagged Regions	8	45	4	35	151	445	151	437

- a. Sequence differences relative to the MT1 reference sample. A common length polymorphism at position 310 was not detected under the conditions used and was excluded from this analysis. However, this polymorphic site has previously been shown to be amenable to screening by oligonucleotide hybridization.
- b. Number of sequence positions called correctly by automated basecaller.
- c. The p^o probe for each target base is either perfectly matched to the target (0) or has 1 mismatch as a result of neighboring polymorphisms.

A total of 12 samples containing 180 substitutions relative to mt1 was analyzed (mt3, mt4, mt5, mt6, ha001, ha002, ha004, ha007, ief002, ief007, ief011 and yr019). Results are summed for all 12 samples. All but two of the 180 substitutions were detected as po intensity differences (one of the exceptions was read correctly by basecalling and automatically flagged for further analysis). In total, only 649 bp (2% of the sequence analyzed) were flagged for further analysis. Basecalling results are broken down separately for unflagged and flagged regions. Fully automated basecalling in unflagged regions had an error rate of 1/127 polymorphisms or 1/29,879 total bp. In contrast, flagged regions, which included 53 (29%) of the substitutions contained 14 false

negatives and 8 false positive. However, we estimated that, on average, 2 to 3 conventional sequencing reactions per sample, or - 30 reactions in total, would resolve the flagged regions, to give a basecalling accuracy in excess of 99.9% for the entire sequence. This represents 8-fold less sequencing than we used to determine the sequences by conventional methods alone and a similar saving in the labor intensive task of sequence checking and editing. In this experiment, samples were prepared and hybridized as described in Fig. 25. In order to provide an independently determined reference sequence, each 2.5 kb PCR amplicon was sequenced on both strands by primer-directed fluorescent chain-terminator cycle sequencing using an ABI373A DNA sequencer, and assembled and manually edited using Sequencer 3.0. Hybridization analysis was also performed on both strands. The analysis presented here assumes that the sequence amplified from genomic DNA is essentially clonal, or at least contains one predominant species, and that its determination by gel-based methods is correct. PCR amplification errors might contribute a maximum of - 0.5% sequence difference, essentially randomly distributed, based on an estimation of - 10^5 fold amplification and known error rates of Taq polymerase. This would not be expected to affect significantly the results of gel based sequencing or hybridization analysis, particularly when analyzing differences from a reference hybridization pattern.

Basecalling was performed using a Bayesian classification algorithm based on variable kernel density estimation. The likelihood of each basecall associated with a set of hybridization intensity values was computed by comparing an unknown set of probes to a set of example cases for which the correct basecall was known. The resulting four likelihoods were then normalized so that they summed to 1. Data from both strands were combined by averaging the values. If the most likely basecall had an average normalized likelihood of greater than 0.6, it was called, otherwise the base was called an ambiguity. The example set was derived from 2 different samples, ib013 and ief005, which have a total of 35 substitutions relative to MT1 of which 19 are shared

with the 12 samples analyzed and 16 are not. Base calling performance was not sensitive to the choice of samples. The hybridization sequence analysis was fully automated, with no user editing. In contrast, conventional sequencing required contig assembly followed by editing, in which > 1% of basecalls were manually corrected.

High Density Oligonucleotide Arrays

Several technologies have been developed to design, synthesize, hybridize and interpret high density oligonucleotide arrays of the type described above. Representative arrays are described in U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and WO 92/10092, each of which is incorporated herein by reference. Often, arrays have a lower limit of 25, 50 or 100 probes and an upper limit of 1,000,000, 100,000, 10,000 or 1,000 probes. A range of lengths of probes may be employed. Preferably, each of the high density oligonucleotide arrays contain 10,000-20,000 oligonucleotide probes (length 10-20 mers) which are used to determine the sequence of a target nucleic acid (RNA or DNA). Frequently, the density of the different oligonucleotides is greater than about 60 different oligonucleotides per 1 cm². The determination of a target sequence is accomplished by carrying out a single hybridization reaction involving all of the probes on the surface of the chip. The following is a brief overview describing the synthesis, array design, sample preparation/fluorescent labeling and base calling features of the DNA chips used in this invention.

Light directed synthesis:

DNA chips use light directed synthesis to build oligonucleotide probes on the surface of the chip (Fodor, et al., Science, 251:767-73 (1991)). This light-directed synthesis combines semiconductor based photolithography and solid phase chemical synthesis. With reference to Fig. 6, the process begins when linkers modified with photochemically removable protecting groups (C) are attached to a solid

substrate, the chip surface. Linkers and phosphoramidites with photolabile protecting groups have been synthesized and are described by Pease, et al., PNAS, 91:11241-11245 (1994). Light is directed through a photolithographic mask to specific areas of the synthesis surface, activating those areas for subsequent chemical coupling. The first of a series of nucleotides (T in Fig. 6) possessing photolabile protecting groups, is incubated with the chip and chemical coupling occurs at those sites which have been illuminated in the preceding step. Light is then directed through a different section of the mask to the next synthesis site and the chemical steps, a defined collection of oligonucleotide probes can be constructed, each having its own unique address on the surface of the chip.

Synthesis of complete and subset-combinatorial arrays:

In a light-directed synthesis, the location and composition of the oligonucleotide products depends upon the pattern of illumination and the order of chemical coupling reagents. Consider the synthesis of a chip containing all possible tetranucleotide oligomers (256 possibilities) (Fig. 7). In cycle 1, mask 1 activates one fourth of the substrate surface (dT). In round 2 of cycle 1, mask 2 activates a different quarter of the substrate for coupling with the second nucleoside (dC). This process is continued to build four regions of mononucleotides. The masks of cycle 2 are perpendicular to those of cycle 1, and each synthesis round generates four new dinucleotides until all 16 possible dinucleotides are made (Fig. 7). The masks of cycle 3 further subdivide the synthesis regions so that each coupling round generates 16 trimers. The subdivision of the substrate is continued through cycle 4 to form all possible 256 tetramers (complete combinatorial array). The successful demonstration of light-directed complete combinatorial array has recently been described (Pease, et al., PNAS, 91:11241-11245 (1994)). It is important to note that any subset of a complete array can be synthesized by modifying the mask patterns used in each cycle and round of synthesis. The complete combinatorial

arrays can be used for applications in which *de novo* sequencing is sought (Fodor, et al., 1993), while a subset of combinatorial arrays can be used for resequencing applications such as will be employed in this application.

5

Sample preparation, and fluorescent labeling of target nucleic acid:

Oligonucleotide arrays hybridized to amplification-generated fluorescently-labeled DNA or RNA and the
10 hybridizations are detected by epi-fluorescence confocal microscopy (Fodor, et al., 1993; Molecular Dynamics, Santa Clara, CA). This process is initiated by the extraction of target nucleic acids from the sample. With reference to Fig. 8, the target nucleic acid (mycobacterium genomic DNA) is
15 amplified by the polymerase chain reaction (PCR) using target gene specific primer pairs containing bacteriophage RNA polymerase promoter sequences (Fig. 8). PCR amplified copies of the target nucleic acid are converted from double stranded (ds) DNA into fluorescently-labeled single stranded (ss) RNA
20 during an *in vitro* transcription (IVT) reaction. Finally, the fluorescein-labeled target gene specific RNA transcripts are fragmented into oligomer length targets under elevated temperature and 30mM Mg⁺⁺. The precise protocols, primers and conditions for sample extraction, amplification, chip
25 hybridization and analysis are described in the Examples. Of course other labelling strategies may be utilized.

Resequencing chips and detection of single base mismatches:

As described earlier, a target gene sequence can be
30 represented on a chip in a series of overlapping oligonucleotide probes arrayed in a tiling strategy (Fig. 9). In such a strategy each base in the target is interrogated by using a collection of 4 oligonucleotide probes which are identical except for the base located at or near the center of
35 the probe. Each of the four probes contains dA, dT, dC or dG at this interrogation position. Of the four oligonucleotide probes the one which is the exact complement will produce the most stable hybrid and thus the strongest fluorescent signal

after post-hybridization washing the DNA chip. Likewise, the next nucleotide in the target sequence can be interrogated with four identical length probes which are the same as the first four except they are offset one nucleotide downstream.

5 The central base of these probes also have all four possible bases. In like fashion, all of the bases of a target sequence can be interrogated using overlapping probes arranged in a tiling strategy. The determination of which of the four possible probes is the complement to target is made by taking
10 the ratio of highest to next highest hybridization signal. If this ratio is greater than 1.2, then a specific determination of the interrogated base can be made. If the highest hybridization signal does not meet this criteria then an ambiguous determination is made based on the IUPAC sequence
15 codes.

The sensitivity of DNA chip probes to detect single base mismatches is illustrated using a 16 step combinatorial synthesis. The photolabile MenPoc-dA and MenPoc-dT were the only nucleotides used during the synthesis of the probes on
20 the chip. The lithographic masks were chosen such that each of 256 octanucleotides were synthesized in four independent locations on a 1.28 x 1.28 cm chip surface. This yielded an array of 1024 octanucleotides each occupying a 400 x 400 μ m synthesis region. Following synthesis and phenoxyacetyl
25 deprotection of the dA amine, the glass substrate was mounted in a thermostatically regulated hybridization cell. The target employed for this experiment was an oligonucleotide 5'-AAAAAAA-fluorescein-3' present in a 1 nM concentration. After 30 minute hybridization and washes with x 1.0 SSPE at 15
30 °C, the chip was scanned using an epi-fluorescent confocal reader. The fluorescent intensities of each of the hybridization events were plotted against the position of the mismatches of the probes on the surface of the chip (Fig. 10). The position zero mismatch (with the perfect complement 3'-
35 TTTTTTTT-5') is the brightest hybridization on the array with the background signal of this array at approximately 220 counts. Mismatch position 1 (at the 3' end of the probe) (3'-ATTTTTTTT-5') is the next highest hybridization. The

resulting "U" shaped curve indicates the relative stability of the mismatches at each position of the probe/target complex. The mismatches at positions 3, 4, 5 and 6 are very destabilizing and the intensities of these hybridizations are approximately 3 fold lower than the perfect match hybridization. It is also noteworthy that the mismatch at position 1 (the point where the octanucleotide is tethered to the chip substrate) is less destabilizing than the corresponding mismatch at position 8 (5' free end). The uniformity of the array synthesis and the target hybridization is reflected in the low variance of the intensities of the four duplicate synthesis sites.

Pattern Recognition Algorithm

Hybridization patterns derived from the oligonucleotide probe arrays can be correlated with the drug resistance phenotype and speciation of the organism using mathematical pattern recognition algorithms, such as tree-structured classification techniques. It is important to note that as the total number of analyzed isolates for each species is increased, it is unlikely that a single and unique core fingerprint will define a mycobacterium species. Rather, it is expected that any particular isolate of a *Mycobacterium* species will have a subset of all possible fingerprints. Identification of the *Mycobacterium* species based on a fingerprint pattern will require a classification analysis built upon a collected database consisting of species specific and shared SNPs and fingerprints.

The goal of identifying an unknown *rpoB* hybridization pattern as coming from one of the *Mycobacterium* species in the data base is a general classification problem. Measurements (sequence and fingerprint data) are made on a collection of samples. Based on these measurements a systematic way is developed to predict the class (species) of each member of the collection. The signal produced by the target at each hybridization site is compared to the signal produced by MT *rpoB*. Based on this comparison one determines

whether or not there is a difference in genotype at the interrogated at that site in the target relative to MT *rpoB*.

Classifier construction is based on past experience.

In systematic classifier construction, past experience is summarized by a *learning sample* (a.k.a. *design or training sample*). This consists of the measurement data on *N* cases observed in the past together with their actual classification. It is intended that the database collected in Phase I will serve as the initial training sample. There are two general types of variables that can appear in the measurement data, *ordered or numerical variables* and *categorical variables*. A variable is called ordered or numerical if its measured values are real numbers. A variable is categorical if it takes values in a finite set not having any natural ordering. In our case, each nucleotide position in the sequence is a variable. So all measurement variables are categorical. The set of measurement variable for a given case is called the *measurement vector*. The *measurement space* is defined as the set of all possible measurement vectors.

The four most commonly used classification procedures are *discriminant analysis*, *kernel density estimation*, and *kth nearest neighbor*, and *tree-structured classification*. Discriminant analysis assumes that all the measurement vectors are distributed multivariate normal, and thus is not set up to handle categorical variables (See Gnanadesikan, R. Methods for statistical data analysis of multivariate observations (Wiley, New York (1977))). Even though kernel density estimation and *kth nearest neighbor* methods make minimal assumptions about the form of the underlying distribution of the measurement vectors, there are still serious limitations common to both methods. They require a definition of a distance measure (metric) among the measurement vectors; performance of these classifiers is sensitive to the choice of the metric. There is no natural or simple way to handle categorical variables. Both are computationally expensive as classifiers because they require that the learning sample be stored, and the distances and classification rule be recomputed for each new undetermined

case. Most seriously, they give very little usable information about the data structure. Kanal, L. (1974) IEEE Trans. Information Theory IT-20:697-722, and Hand, D.J. Discrimination and Classification (Wiley, Chichester (1981)),
5 give surveys of the literature on these methods.

Tree-structured classification is a recursive and iterative procedure. It proceeds by repeated splits of subsets of the measurement space into two descendant subsets or nodes, beginning with the full measurement space. The
10 fundamental approach is to select each split of a node so that the data in each of the descendant nodes are "purer" than the data in the parent node. A node impurity measure is defined such that it is largest when all classes are equally mixed together in that node, and smallest when the node contains
15 only one class. The sequence of splits is determined such that at each candidate node all possible splits are examined and the split that produces the largest decrease in the impurity is selected. The terminal nodes form a partition of the measurement space. Each terminal node is designated by a
20 class assignment based on the observed proportions of the classes in that partition. (Usually, the assignment is the class with the highest proportion.) There may be two or more terminal nodes with the same class assignment. The partition corresponding to that class is obtained by putting together
25 all terminal nodes corresponding to the same class. The tree classifier predicts a class for a given measurement vector in the following way: From the definition of the first split, it is determined whether the measurement vector goes to the right or to the left. This is repeated until the case moves into a
30 terminal node. The predicted class is then given by the class assignment of that terminal node.

The optimal size of a classification tree is determined in the following manner: continue the splitting until all terminal nodes are very small, resulting in a large
35 tree. This large tree is then selectively pruned upward, and thus producing a decreasing sequence of subtrees. Finally, use cross-validation or test-sample estimates to choose from the sequence of subtrees that subtree having the lowest

estimated misclassification rate. The tree-structured classification methodology is covered in detail in Breiman et. al. Classification and Regression Trees, (Wadsworth International Group, Belmont, California (1984)).

5 The tree-structured approach is a powerful and flexible classification tool. It can handle both numerical and categorical variables in a simple and natural way. The final classification has a simple form that can be efficiently used to classify new data. It does automatic stepwise
10 variable selection and complexity reduction. It provides both the classification and the estimate of the misclassification probability for a new case. The output of the tree procedure gives easily understood and interpreted information regarding the predictive structure of the data.

15 Example of a Tree-Structured Classifier

 Figure 11 displays a hypothetical six-class tree (numbers under boxes). The boxes represent nodes. Node t_1 contains the whole measurement space and is called the root node. Nodes t_2 and t_3 are disjoint with t_1 being the union of
20 t_2 and t_3 . Similarly t_4 and t_5 are disjoint and t_2 is the union of t_4 and t_5 . Those nodes that are not split, in this case, t_6 , t_8 , t_{10} , t_{11} , t_{12} , t_{14} , t_{15} , t_{16} , and t_{17} are called *terminal nodes*. The numbers beneath the terminal nodes are
25 the *class assignments* or *class labels* for this particular classifier.

 Let x be a measurement vector (in our case a DNA sequence of length K); $x = (x_1, x_2, \dots, x_K)$. The splits are formed by setting conditions on the coordinates of x . For
30 example Split 1 of t_1 into t_2 and t_3 could be of the form: t_2 is the set of measurement vectors such that $x_{10} = \{A, C\}$ and t_3 is the set of measurement vectors such that $x_{10} = \{G, T\}$.

 This classifier predicts a class for the measurement vector x in this way: From the definition of the first split,
35 it is determined whether x goes into t_2 or t_3 . For example, if the above definition for Split 1 is used, x goes into t_2 if the 10th nucleotide in that sequence is either A or C, otherwise, x goes into t_3 . If x goes into t_2 , then from the

definition of Split 2, it is determined whether *x* goes into *t*₄ or *t*₅, and so on. When *x* finally moves into a terminal node, its predicted class is given by the class label attached to that terminal node.

5

EXAMPLES

Mycobacterium tuberculosis *rpoB* chip

A high density oligonucleotide array (*Mtb rpoB* chip) has been synthesized and tested in preliminary experiments.

10 The chip has been synthesized using 2 lengths of oligonucleotides (18 and 20 mers) with the interrogation position located at bases 9 and 10 (sense and antisense probes) and 10 and 11 (sense and antisense probes). The *Mtb rpoB* chip was used initially to genotype 15 *M. tuberculosis* clinical isolates which were previously determined to be RIF sensitive. Figure 12 is an image of the *Mtb rpoB* chip analysis of the 700 bp region of the *rpoB* gene from one of these isolates. Oligonucleotide primer sequences, PCR amplification, in vitro transcription and hybridization to the chip conditions were as follows.

20 Chromosomal DNA from *M. tuberculosis* was isolated by suspending one colony in 100 μ l of ddH₂O, boiling for 10 minutes and briefly centrifuging to separate the DNA solution from cellular debris. The chromosome DNA was then diluted 25 1:10 in ddH₂O. A 705 bp *rpoB* fragment was amplified in a 100 μ l reaction volume containing each dNTP at 200 μ M, each primer at 0.2 μ M, 2.5 U of Taq-polymerase (BM, Indianapolis, IN), 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂. The amplification was carried out in a model 9600 thermocycler (Perkin Elmer 30 Cetus). To amplify the 705 bp fragment using primers *rpoB*-4 (CTC GGA ATT AAC CCT CAC TAA AGG GAC CCA GGA CGT GGA GGC GAT CAC ACC GCA) (SEQ ID NO:1) and *rpoB*-7 (TAA TAC GAC TCA CTA TAG GGA GAC GTC GCC GCG TCG ATC GCC GCG C) (SEQ ID NO:2) with incorporated T3 and T7 sequences 5 min 95°C, 35 cycles of 1 35 min 95°C, 30 sec 68°C and 2 min at 72°, followed by 10 min of 72° were used. The PCR amplicon was then purified using Amicon Microcon 100 columns. In vitro transcription of approximately 50 ng amplicon was performed in a reaction

volume of 20 μ l, containing 1.25 mM rNTPs, 10 mM DTT, 125 μ M F-CTP, 20 U RNase inhibitor, 40 mM Tris-HCl (pH7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 20 U T3/T7 RNA-polymerase for 90 min at 37°C. The RNA was then purified with Amicon
5 Microcon 100 columns and quantitated using a spectrophotometer. Fragmentation was carried out in 30 mM $MgCl_2$ at 95°C for 30 min. A 20 nM RNA solution in 6xSSPE, 20% formamide, 0.005% Triton, 0.5 nM control oligo was heated to 68°C for 10 min, then placed on ice for 5 min and hybridized
10 to the Mtb *rpoB* chip for 30 min at 22°C in the Affymetrix Fluidics Station. The post-hybridization wash was performed with 1xSSPE, 20% formamide, 0.005% Triton x 100 in the Affymetrix Fluidics Station (Affymetrix, Santa Clara, CA), using 12 wash cycles with 2 fills and drains per cycle,
15 followed by a wash with 6xSSPE, 0.005% Triton, 2 cycles with 2 fills and drains per cycle. The chip was then scanned on a Molecular Dynamics scanner (Molecular Dynamics, Santa Clara, CA) at 22°C with a resolution of 11.25 pixels/ μ m.

As noted, 20% formamide was used in both
20 hybridization and post-hybridization wash steps since the 700 bp amplicon is 67.7% G:C rich with a 18 bp region which is 73.3% G:C rich. The results from the analysis by the Mtb *rpoB* chip indicated that there were no polymorphisms at any base of the 700 bp for any of the 15 *M. tuberculosis* RIF sensitive
25 isolates analyzed. This result was confirmed by conventional dideoxynucleotide sequencing. Thus, both methods were 100% concordant in the analysis of 10,500 nucleotides of total sequence.

30 Detection of mutations conferring RIF resistance

Four pre-resistant/post-resistant RIF isolates were screened in a blinded fashion. These were analyzed using this first generation Mtb *rpoB* chip. Table 2 summarizes the results of the chip analysis. Of the 4 pair isolates, three
35 pair were observed to have one member of the pair which possessed mutations in the 81 bp region (all other nucleotides were wild type), with the companion isolates displaying no such mutations. Interestingly, the fourth pair

(001415/001417) contained no mutation at any nucleotide of the 700 bp surveyed, although isolate 001417 was characterized as RIF resistant by culture assay. Since 10% of RIF resistant *M. tuberculosis* isolates have no mutation in the 81 bp region of *rpoB*, this isolate may be resistant because of a mutation in the portion of *rpoB* not analyzed by the chip or because of a mutation in some other gene which controls uptake, metabolism or drug binding. The sequences derived using the chip for all 8 isolates were confirmed using dideoxynucleotide sequencing. An additional 4 RIF resistant isolates were also screened. Mutations only within the 81 bp region were detected for each of these isolates by the *Mtb rpoB* chip and confirmed by dideoxynucleotide sequencing. A total of 25 *M. tuberculosis* isolates were analyzed. Seven of these were rifampicin resistant and had the mutations shown in Table 2. Other than the mutations shown in Table 2, there were no polymorphisms in any of the 25 isolates.

Table 2

RIF Sensitive and Resistant *M. tuberculosis* Clinical Samples Analyzed
by *Mtb rpoB* Chip and Confirmed by Did

Sample	Amino Acid ¹	Nucleotide ¹	Phenotypic Resistance ¹
M0404A	None	None	No
000936	S456L	TCG->TTG	Yes
00145	None	None	No
001417	None	None	Yes
000914	None	None	No
001231	S456L	TCG->TTG	Yes
001587	H451D	CAC->GAC	Yes
SM2341	None	None	No
3407	H451Y	CAC->TAC	Yes
978	H451L	CAC->CTC	Yes
3553	S456L	TCG->TTG	Yes
3466	S447L	TCG->TTG	Yes

¹. Amino acid and nucleotide numbering system employs sequencing derived by Miller, et. al., (1993).
². Resistance was determined using relative/proportion method of Small et al., "Tuberculosis: Pathogenesis, Protection, and Control" pp. 569-586 (1994) (Am. Soc. Microbiol., ed. B.R. Bloom)

Single Nucleotide Polymorphisms and Hybridization Pattern (Fingerprint) Database for Non-Tuberculosis *Mycobacteria*

The first steps in assembling a database (consisting of SNPs and chip based hybridization fingerprints) capable of identifying mycobacteria species were taken with the analysis of 7 clinically important *Mycobacteria* species: *M. gordonae*, *M. chelonae*, *M. kansasii*, *M. scrofulaceum*, *M. avium*, *M. intracellulare* and *M. xenopi*. As a first step, the 700 bp region of the *rpoB* gene from one isolate from each of these species was sequenced using dideoxynucleotide methodology. Nucleotide (60-71) and amino acid (5-8) differences were compared to *M. tuberculosis* within the 700 bp region for each of these mycobacteria species (Table 3). Two types of single nucleotide polymorphisms (SNP) were noted: species specific (unique) and shared. The SNPs which were shared with at least 3 other non-tuberculosis mycobacteria are numerous and scattered throughout the 700 bps analyzed. Fig. 13 illustrates the location of these shared SNPs. The species specific SNPs are, however, considerably fewer. Figure 14 depicts the location and nature of the SNPs for each of the 7 species analyzed based on one isolate for each species.

Table 3

Comparison of Polymorphisms in *Mycobacteria* Species
Number of polymorphic

Strain	Nucleotides		AA - changes	
	Total	Unique	Total	Unique
<i>M. gordonae</i>	71	19	5	1
<i>M. chelonae</i>	62	1	5	1
<i>M. avium</i>	60	10	5	0
<i>M. kansasii</i>	67	17	6	4
<i>M. scrofulaceum</i>	63	1	6	1
<i>M. xenopi</i>	72	27	8	7
<i>M. intracellulare</i>	60	7	6	2

When fluorescently-labeled RNA amplicons from each of the 7 mycobacteria species were hybridized to the *Mtb rpoB* chip, the image of the hybridization is considerably different then when an amplicon for *M. tuberculosis* was hybridized (Fig. 15A). The differences in the hybridization patterns can be

represented as a bar-code-like fingerprint (Fig. 15B). Each line on the fingerprint represents a hybridization difference as compared to when wild type *M. tuberculosis* is hybridized. These differences are attributable to the species specific and shared polymorphisms identified by the dideoxynucleotide sequencing analysis (Table 3). For any individual *Mycobacteria* species, only some of the differences depicted by an individual line of the fingerprint is identifiable as a specific base pair difference. The remainder of the lines of the fingerprint can be characterized only as being different than if a *M. tuberculosis* sample was hybridized. These undefined differences are usually caused when multiple polymorphisms occur in close proximity or they are the result of the destabilization of the hybridization of neighboring probes due to the presence of unique or shared polymorphisms. In such clustered polymorphism cases, there are multiple mismatches within individual probes interrogating nucleotides in a region. Hybridization results involving such probes are unstable leading to ambiguous, wrong or no calls. Thus, a full fingerprint pattern is composed of identified (unique or shared polymorphisms) and unidentified (clustered polymorphisms leading to no base calling determination) differences. An average of 27.7% of the 700 bps interrogated are viewed as different than if a *M. tuberculosis* target were hybridized to the chip (Table 4). In other words, for every base identified by ABI sequencing as being polymorphic, the chip sees three bases as different from Mtb. This is due to, in addition, each of the two bases flanking the base identified by ABI as polymorphic also being viewed as different by the chip, because of destabilization of hybridization at these sites.

Table 4

Differences of Fingerprint Patterns Among
Mycobacteria Species Compared to *M. tuberculosis*

		Nucleotide Differences ¹	% Differences ²
5	<i>M. gordonae</i>	188	26.7
	<i>M. chelonae</i>	208	29.5
	<i>M. avium</i>	152	21.5
	<i>M. kansasii</i>	216	30.6
	<i>M. scrofulaceum</i>	213	30.2
10	<i>M. xenopi</i>	229	32.4
	<i>M. intracellulare</i>	164	23.2

¹ The nucleotide differences are composed of identified differences compared to the *M. tuberculosis* sequence (species specific and shared polymorphisms) and unidentified differences (caused by clustered polymorphisms).

² % differences are based on a total of 700 bp analyzed by *Mtb rpoB* gene on the chip.

15 Since the database for each of the non-tuberculosis mycobacteria was the result of analysis of only a single isolate for each *Mycobacteria* species, the variation of fingerprint patterns that would be observed among multiple
 20 isolates of a single *Mycobacteria* species was explored. Consequently, the *rpoB* gene from 10 isolates of *M. gordonae* were analyzed by the *Mtb rpoB* chip. Figure 16 presents the images of the sense strand hybridization. Below each chip
 25 image is the hybridization fingerprint computed after analysis of both strands. The shared differences among the 11 (10 new and 1 original) isolates analyzed are shown below (Table 5). From this analysis a core (consensus) fingerprint pattern for *M. gordonae* was derived (Figure 16). A similar core
 30 fingerprint has been derived for eight other *Mycobacterium* species, thus allowing identification of those species. It will be apparent that the techniques described above can be used to assemble a database of species-specific and shared polymorphisms which can be used to derive fingerprints for other species.

35 It is important to note that as the total number of analyzed isolates for each species is increased, it is unlikely that a single and unique core fingerprint will define a *Mycobacterium* species. Rather, it is expected that any particular isolate of a *Mycobacterium* species will have a
 40 subset of all possible fingerprints. Identification of the

Mycobacterium species based on a fingerprint pattern will require a classification analysis, as described earlier, using the tree-based classification algorithms built upon a collected database consisting of species specific and shared SNPs and fingerprints.

Table 5

Percent Shared Differences Among *Gordoniae* Clinicals

	golz	gord	gordjd	gordib	gordig	gordil	gordmb	gordow	gordrb	gordwn	gorm
golz	0										
gord	18.7	0									
gordjd	16.7	17.4	0								
gordib	19.3	22.7	18.3	0							
gordig	23.7	18.7	17.3	19.6	0						
gordil	18.3	14.6	15.2	15.5	17.9	0					
gordmb	19.7	22.4	18.4	24.3	19.7	15.3	0				
gordow	19.6	16.3	16.6	17.0	19.6	18.7	16.7	0			
gordrb	20.6	23.3	18.9	24.5	20.6	16.2	25.0	18.4	0		
gordwn	17.6	17.4	16.9	17.9	16.9	15.2	18.0	16.2	19.0	0	
gorm	20.1	22.7	19.1	24.0	20.9	16.2	24.3	17.6	25.0	18.2	0

Human Mitochondrial DNA Chip (MT1)

Fluorescein labelled target RNA was synthesized and fragmented, and the transcription mixture diluted 20-fold in 6xSSPE, 0.05% Triton X-100, to give approximately 1 to 10 nMRNA (estimated prior to fragmentation). Hybridization was carried out for 30 min at RT. The chip was washed for a total of 5 to 10 minutes in several changes of 6xSSPE, 0.005% Triton X-100, and scanned.

Fig. 23A shows the design of the tiled array on the MTI chip. Each position in the target sequence (upper case) is interrogated by a set of 4 probes on the chip (lower case), identical except at a single position, termed the interrogation base, which is either A, C, G, or T. The target will be perfectly complementary to one of the 4 probes, but mismatched with the others. As illustrated in Fig. 10, the perfect complement gives a more intense hybridization signal than do the mismatches. Each of the lower three probes represents a 4 probe set, with n = A, C, G, or T. By tiling the sets across the sequence in single base increments as

shown, a nucleic acid target of length N can be scanned for mutations with a tiled array containing 4N probes. (B)

Hybridization to a tiled array and detection of a point mutation. The array shown was designed to the MT1 target sequence. When hybridized to MT1 (upper panel), one probe in each set of 4 in a column is perfectly matched to the target, while the other three contain a single base mismatch. The interrogation base used in each row of probes is indicated on the left of the image. The target sequence can be read 5' to 3' from left to right as the complement of the interrogation base with the brightest signal. Hybridization to MT2 (lower panel), which differs from MT1 in this region by a T -> C transition, affects the probe sets differently. At the location of the polymorphism, the G probe is now a perfect match to MT2, with the other three probes having single base mismatches. (A*, C*, G*, T* counts). However, at flanking positions, the probes have either single or double base mismatches, since the MT2 substitution now occurs away from the interrogation position. The location of the mismatch is illustrated in the probe schematic by red circles.

Detection of base differences of base differences between a sample and reference sequence in 2.5 kb by comparison of scaled $p^{0.9}$ hybridization intensity patterns between a sample and a reference target

In this study, each 2.5 kb target sequence was PCR amplified directly from genomic DNA using the primer pair L14675-T3 (5'aattaaccctcactaaagggATTCTCGCACGGACTACAAC) (SEQ ID NO:7) and H667-T7, transcribed to give RNA targets labelled with fluorescein or biotin, pooled and fragmented as described. In the experiments shown the MT1 reference target was biotin labelled and the sample target fluorescein labelled. Targets were diluted 180 fold from the transcription reaction to a final concentration of - 100 to 1000 pM in 3 M TMACl, 10 MM Tris.Cl pH 8.0, 1mM EDTA, 0.005% Triton X-100, and 0.2 nM of a control oligonucleotide, 51 fluorescein-CTGAACGGTAGCATCTTGAC (SEQ ID NO:8). (We found that the G-rich H strand target hybridized poorly in 1 M NaCl, but

hybridized well in 3 M tetramethyl ammonium chloride, whereas the L strand hybridized well in either solution).

Hybridization was carried out in packaged chips. Samples were denatured at 95°C for 5 min. chilled on ice for 5 min. and equilibrated to 37°C. A volume of 180 μ l of hybridization solution was then added to the flow cell and the chip incubated at 37°C for 3 h with rotation at 60 rpm on a laboratory rotisserie. Following hybridization, the chip was washed 6 times at RT with 6xSSPE, 0.005% Triton X-100. A solution of 2 μ g/ml phycoerythrin-conjugated streptavidin in 6xSSPE, 0.005% Triton X-100, was added, and incubation continued at RT for 5 min. The chip was washed again, and scanned at a resolution of - 74 pixels per probe cell. Two scans were collected, one using a 530 DF 25 nm bandpass filter, and the second using a 560 nm longpass filter. Signals were deconvoluted to remove spectral overlap and average counts per cell determined. The sample probe intensities were scaled to the reference intensities as follows: a histogram of the base-10 logarithm of the intensity ratios for each pair of probes was constructed. The histogram had a mesh size of 0.01, and was smoothed by replacing the value at each point with the average number of counts over a five-point window centered at that point. The highest value in the histogram was located, and the resulting intensity ratio was taken to be the most probable calibration coefficient.

The data are shown in Fig. 25 for L strand targets hybridized to H strand probes, from a portion of hypervariable region I in the mitochondrial control region. Numbering is conventional. In each plot, the reference target intensities are shown in red and the sample in blue. The reference, MT1, is a perfect match to the p⁰ probes. Fig.25A - Comparison of ief007 to MT1. There is a single base difference between the two target sequences, located at position 16,223 (MT1 C: ief007 T). This results in a "footprint" spanning - 20 positions, 11 to the left and 8 to the right of position 16,223, in which the ief007 p⁰ intensities are decreased by a factor of more than 10 fold relative to the MT1 intensities.

The size and location of the footprint are consistent with a single base mismatch affecting hybridization to $p^{20,9}$ probes. The theoretical footprint location is indicated by the grey bar, and the location of the polymorphism is shown by a vertical black line within the bar. The size of a footprint changes with probe length, and its relative position with interrogation position (not shown). Because the sample and reference targets are in competition, the MT1 signal in a footprint region actually increases as a proportion of total signal in each probe cell, because the mismatched sample target no longer competes effectively for probe sites.

Fig.25B - Comparison of ha001 to MT1. The ha001 target has 4 polymorphisms relative to MT1. The p^0 intensity pattern clearly shows two regions of difference between the targets. Furthermore, it can easily be seen that each region contains ≥ 2 differences, because in both cases the footprints are longer than 20 positions, and therefore are too extensive to be explained by a single base difference. The effect of competition can be seen by comparing the MT1 intensities in the ief007 and ha001 experiments: the relative intensities of MT1 are greater in panel B where ha001 contains no mismatches but ief007 does not. Fig. 25C - The ha004 sample has multiple differences to MT1, resulting in a complex pattern extending over most of the region shown. Thus, differences are clearly detected, even though basecalling might be compromised using only the 4N tiling array. Even when patterns are highly complex, samples can be compared and matched. For example, if the ha004 sample is compared to ha004 as a reference in the same experiment, the p^0 pattern indicates a match, even though the effect of multiple mismatches might compromise direct sequence reading (not shown).

Detection of a 2 bp-deletion

Experimental details are as above. The results are shown in Figure 26. Figure 26A shows that although the 4N array was not designed to detect length polymorphisms, this common 2 bp length polymorphism located at 514-523 in the MT DNA was easily detected by the presence of a p^0 intensity

footprint. Figure 26B shows target-specific effects on hybridization. A 2 bp difference between the MT1 reference (GG) and the ha002 target (AA) is associated with a complex footprint pattern: the p₀ signals of the mismatched ha002 hybrids are up to 10-fold higher than those of the perfectly attached mt1 hybrids in a region extending leftwards from position 16,381. Both samples were hybridized simultaneously to the same array. In addition, the effect extends beyond the probes directly affected by the mismatches. Therefore, we conclude that the difference is due to changes in target secondary structure. Increased accessibility of the target as a result of disruption of base pairing between inverted repeats (shaded on the diagram) could explain the increase in the sample versus the reference p₀ signals.

Hybridization of a 16.3 kb of mitochondrial target sequence to the whole genome chip

Figure 27A shows an image of the array, actual size hybridized to L strand target sample. The 1.28 x 1.28 cm, p^{20,9} tiled array contains a total of 134,688 probes, each synthesized in a 35 x 35 micron cell. The number of probes is sufficient to represent the 16.6 kb genome twice over. The array has the capacity for sense and antisense coding. The 16,569 bp map of the genome is shown and the H strand origin of replication (OH)' located in the control region, is indicated. Figure 27B - A portion of the hybridization pattern is shown, magnified. The scale is indicated by the bar on the left hand side. Most of the array can be read directly. The image, which was generated by the galvanometer scanner detection system in under 2 minutes, was collected at -3 micron, 16 bit pixel resolution, providing - 100 pixels of intensity data for each probe cell in the array. Fluorescence was detected through a 581 Df 52 nm. bandpass filter. Figure 27C - The ability of the array to detect and read single base differences in a 16.3 kb sample is illustrated. Two different target sequences were hybridized in parallel to different chips. The hybridization patterns are compared for four different positions in the sequence. The top panel of each

pair shows the hybridization of a the MT3 target, which matches the chip po sequence at these positions. The lower panel shows the pattern generated by a sample from a patient with Leber's Hereditary Optic Neuropathy (LHON). Three
5 pathogenic mutations are implicated, LHON3460, LHON4216, and LHON13708. All three are clearly detected. For comparison, the third panel in the set shows a region that is identical in both samples, around position 11,778.

The pattern matching techniques described above also
10 provide a method of determining whether the nucleic acid sequence of a biological sample is homozygous or heterozygous for a particular allele, i.e., to identify the presence of a polymorphism in the nucleic acid sequence at a particular position. In this regard, polymorphisms can be identified in
15 both coding and noncoding sections of the sample nucleic acid, i.e., in exons or introns. This is of value, for example, in identifying whether a genetically linked disease is present. Of course, it will be recognized that any genetically related condition, i.e., other than those thought as "diseases" can be
20 identified by such a method.

Fig. 34 shows a computer-implemented flowchart of a method of identifying the presence of a polymorphism in a nucleic acid sequence from a patient sample. The flowcharts described herein are for illustration purposes and not
25 limitation. For example, for simplicity Fig. 34 describes analyzing one base position at a time to detect polymorphisms. However, each step may also be performed for the entire nucleic acid sequences and/or some steps may be combined.

At step 200, the system selects a base position in
30 the nucleic acid sequence from the patient sample. The system determines the difference between the hybridization intensities of the nucleic acid sequence from the patient sample and a corresponding nucleic acid sequence from a wild type sample to an array of reference nucleic acid probes at
35 step 202. Although the reference nucleic acid probes may perfectly complementary to the wild type sample, this is not required.

The system derives or calculates a ratio of the difference determined at step 202 to the hybridization intensity of the nucleic acid sequence from the wild type sample. The ratio is derived at step 204 and it indicates how close the hybridization intensities for the nucleic acid sequences from the patient wild type samples are to being the same.

An assigned value is utilized to determine if the ratio indicates that there is a polymorphism at the base position. The assigned value may be user specified and at step 206, the system compares the ratio to the assigned value. If the ratio is greater than the assigned value, the system identifies the presence of a polymorphism at the base position of the nucleic acid sequence from the patient sample. At step 210, the system determines if there is a next base position to analyze.

By way of example and not limitation, one can screen nucleic acid samples from a cancer patient to determine whether the DNA repair genes MSH or MLHI are mutated. This is done by comparing the hybridization pattern of patient DNA from the appropriate region to the hybridization pattern of DNA from the same region of a healthy (i.e., wild-type) sample. Figures 28-31 show such comparisons of patient DNA samples from heterozygous MSH2, MLH1, MSH2 and p53 genes and their corresponding wild type genes. The screening can be against any reference sequence immobilized on the chip, though as described earlier it will be advantageous to use a chip in which the reference sequence is complementary to the wild type sequence. The hybridization intensities corresponding to each base position is determined for each sample as described earlier. One then determines the difference between the intensities for the patient sample and the wild type sample at each base position and compares that to the wild type intensity at that base position. This ratio can vary between one and zero, being zero if the wild type and patient sequences are identical in this region (since hybridization will be identical for both samples) and approaching one if there is a complete mismatch, i.e., no hybridization at all

between patient sample and the reference sequence in that region. If this ratio is greater than an assigned value, it indicates a polymorphism at that particular base position. Typically, this assigned value is set at about 0.5, preferably 0.6. Positions at which such polymorphisms are present can be identified by plotting this ratio versus the corresponding base position for all positions where the ratio is greater than about 0.25. If the ratio is less than 0.25, this is considered to be statistical noise. Typically such plots will show a spike, with a maximum ratio of about 0.5, centered approximately around the site of the polymorphism. The plots are made with variables derived as follows:

Y axis: $y = (\text{WT intensity} - \text{PS intensity}) / \text{WT intensity}$

X axis: $x = \text{base position}$

where WT = wild type and PS = patient sample

The technique has been refined further to provide a higher level of accuracy by determining hybridization intensities from both the sense and antisense strands of the DNA sample and requiring that the spike occur in both strands at the same respective complementary positions. The probes on the chip are typically 10-20 mers and therefore create a "footprint" as one tiles through the position where the polymorphism is present, i.e., there will be a difference between the hybridization intensities of the patient sequence and the wild type sequence in this region. As a result, an even higher level of confidence that a polymorphism occurs at a particular base position is obtained by requiring that the hybridization intensity ratio derived above be greater than the assigned value, 0.5 in this example, for at least two adjacent base positions, preferably three adjacent positions.

Mismatch Detection by Tiled Arrays

In this example, a reference target T0 and three mutant targets T1, T2, and T3 are provided. T1 has a substitution at position 11, T2 at position 9, and T3 at positions 9 and 11. In writing the mutant sequences, the

position of the substitution is noted by S. These sequences are depicted in Table 6.

Table 6

Substitutions in Mutant Sequences

	T0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	T1											S										
	T2								S													
10	T3							S		S												

Each of these targets is hybridized with a DNA chip containing a tiled arrays of probes. For simplicity, a p⁷₄ chip is described herein. The superscript 7 indicates that the chip contains a tiled array of 7-mer probes that are perfectly or partially complementary to the reference target. The subscript 4 denotes the interrogation position, such that the nucleotide at position 4 of each 7-mer is varied (A, T, G, or C in four different synthesis cells).

The number of target-probe mismatches is given in Table 7 below. The top row gives the number of mismatches for the best-match case (i.e., for the most complementary probe of each set of four) and the second row gives the number of mismatches for the other three probes in the set.

Table 7

Target Probe Mismatches

	T0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
30	versus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	versus	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0
35	T1	1	1	1	1	1	1	1	2	2	2	1	2	2	2	1	1	1	1	1	1	1
	versus	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
	T2	1	1	1	1	2	2	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1
40	versus	0	0	0	0	1	1	1	1	2	2	1	1	1	1	0	0	0	0	0	0	0
	T3	1	1	1	1	2	2	2	2	3	3	2	2	2	2	1	1	1	1	1	1	1

Shown below in Table 8 are the number of mismatches in the best-match case for the hybridization of p^7 , with a series of targets containing two substitutions at different separations:

Table 8
Best Match Hybridization

	TO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
10									S	S												
						1	2	2	1	1	2	2	1									
15									S		S											
						1	1	2	1	2	1	2	1	1								
									S			S										
						1	1	1	1	2	2	1	1	1	1							
20									S				S									
						1	1	1	0	2	2	2	0	1	1	1						
									S					S								
25						1	1	1	0	1	2	2	1	0	1	1	1					
									S						S							
						1	1	1	0	1	1	2	1	1	0	1	1	1				

When a target is hybridized to a 7-mer chip P^7 containing the tiled reference sequence, the number of mismatches in this case is the same as that given by the best-match case above except for an additional mismatch at each substitution position (See Table 9). For T3, for example, P^7 has one more mismatch at positions 8 and 11, where T3 has substitutions, than does P^4 .

Table 9
Mismatches at Substitution Position

p' S S

versus 0 0 0 0 1 1 1 2 2 2 2 1 1 1 0 0 0 0 0 0

T3

Thus, to generalize, for a P^n_j chip, and a target containing a substitution at position a, the best-match set of probes will contain 1 mismatch from a-1 to a-k, 0 at a, and 1 mismatch from a+1 to a+m, where $k=j-1$ and $m=n-j$. For P^7_4 , $k=3$ and $m=3$, and so the 1 mismatch zone is from a-1 to a-3, and from a+1 to a+3, with no mismatch at a, the interrogation position:

The effects of multiple substitutions are additive. Thus, for example, using a P^{16}_{10} chip and a target containing substitutions at positions a and b, where k is 9 and m is 6, the effect of the substitution at a is to give 1 mismatch from a-1 to a-9, and from a+1 to a+6. The substitution at b will give 1 mismatch from b-1 to b-9, and from b+1 to b+6. If a and b are at positions 100 and 108, their effect is the sum, as shown in Table 10.

Table 10
Effects of Multiple Substitutes

	<u>Position</u>	<u>Mismatches</u>
	91-98	1
	99	2
	100	1
	101-106	2
	107	1
	108	0
	109-114	1

Thus, given a hybridization pattern, the location of substitution mutations can be done as follows.

(a) The first step is to hybridize a P^n_j chip with the reference target, and another P^n_j chip with the unknown target. Alternatively, a single P^n_j chip could be hybridized with a mixture of differently labeled reference and unknown targets (e.g., a red-fluorescent reference target and a green-fluorescent unknown target). By using a pair of chips, or a pair of suitably labeled targets, one can readily identify

probes that contain mismatches and distinguish between 0, 1, 2 and a larger number of mismatches.

(b) A substitution at location a is identified by the presence of a 1-mismatch zone that is n -residues long except for a 0-mismatch cell at residue a . The probe giving the highest intensity at residue a identifies the substitution.

(c) A "quiet zone" (i.e., where the unknown target exhibits 1 or more mismatches) that is longer than n must contain at least two substitutions (the effects of insertions and deletions are considered below). The differences between P^n_j and P^n reveal the sites of the substitutions. Again, the probe of P^n_j giving the highest intensity at each of these sites identifies the substitution. An example is provided in Table 11 below.

Table 11

	5	6	7	8	9	10	11	12	13	14	15	16
Target				S					S			
P^7	1	1	1	1	1	2	2	1	1	1	1	1
P^7_4	1	1	1	0	1	2	2	1	0	1	1	1
$P^7 - P^7_4$	0	0	0	1	0	0	0	0	1	0	0	0

$P^7 - P^7_4$, the difference between the tiled reference sequence and the best-case match of the tiled array, exhibits 1's at positions 8 and 13 and 0's elsewhere, showing that substitutions have occurred at these two positions. Their identity is established by seeing which of the four nucleotides at these interrogation positions has the highest intensity.

(d) Further information can be obtained by hybridizing a generic chip, such as one containing all 10-mers of DNA, with both the reference target and the unknown target. The difference in hybridization patterns identifies probes that span mutation sites.

Identifying Species Utilizing Generic Probe Arrays

It has been determined that generic high density DNA probe arrays may be utilized to identify species of isolates. By "generic" it is meant that the probe array was not specifically designed to identify species within the genus of interest. For example, a probe array that includes all nucleic acid probes ten nucleotides in length would be a generic probe array. Additionally, a probe array for an entirely different purpose may be utilized as a generic probe array. Thus, a probe array for detecting mutations in HIV may be utilized to identify species in *Mycobacterium*.

Given multiple isolates that one wants to determine the species of each isolate, the isolates are first hybridized with the generic probe array to obtain hybridization intensities as described above. Typically, the hybridization intensities will then be normalized across the isolates. It has been determined that analyzing each hybridization intensity from the experiments may not be computationally feasible, or at least economically feasible. Accordingly, the invention reduces the number of variables to analyze.

In one embodiment, for each probe in the generic probe array, the mean and variance for the hybridization intensities across the isolates is calculated. The probes that demonstrate the most variance are then selected and the corresponding hybridization intensities are utilized to cluster the isolates into species. Thus the invention utilizes the hybridization intensities from probes that have the most varying hybridization intensities. One may first specify a number of probes which one believes could be processed by the equipment available. Then, the invention would select the hybridization intensities from that number of probes which will provide the most discriminating information.

This process may be generally utilized to assign groups to multiple isolates, where the groups are species, subspecies, phenotypes, genotypes, and the like. For illustration purposes, the following will describe an embodiment that identifies the species of isolates.

Fig. 36 shows a computer-implemented flowchart of a method of identifying species to which organisms belong. At step 400, hybridization intensities indicating hybridization affinity between multiple isolates and a generic probe array are input. Optionally, the hybridization intensities are then standardized or normalized at step 402 to reduce the variability between the experiments. For example, the hybridization intensities may be standardized to a common mean and variance by Z-score analysis or normalization.

The system selects hybridization intensities that have the most variance across the isolates at step 404. Determining which hybridization intensities vary the most may be done any number of ways including calculating a mean and variance. As an example, the number of hybridization intensities to analyze may be reduced from 10,000 to 20 which drastically reduces the computational time required to analyze the hybridization intensities.

At step 406, the species of each of the multiple isolates is determined according to the selected hybridization intensities. Clustering algorithms may be utilized to cluster the isolates into species. As an example, Principal Components analysis and Variable Clustering analysis may be utilized. The purpose of clustering is to place the isolates into groups or clusters suggested by the data, not defined a priori, such that isolates in a give cluster tend to be similar and isolates in different clusters tend to be dissimilar. Thus, no a prior classification is required.

Isolates of *Mycobacterium* have been analyzed and Fig. 37 shows a hierarchical clustering of these isolates. The height of the cluster represents the average distance between the clusters.

The foregoing invention has been described in some detail by way of illustration and example, for purposes of clarity and understanding. It will be obvious to one of skill in the art that changes and modifications may be practiced within the scope of the appended claims. Therefore, it is to be understood that the above description is intended to be illustrative and not restrictive. The scope of the invention

should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the following appended claims, along with the full scope of equivalents to which such claims are entitled.

5

All patents, patent applications and publications cited in this application are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual patent, patent application or publication were so individually denoted.

WHAT IS CLAIMED IS:

1 1. A method for identifying a genotype of a first
2 organism, comprising:

3 (a) providing an array of oligonucleotides at known
4 locations on a substrate, said array comprising probes
5 complementary to reference DNA or RNA sequences from a second
6 organism;

7 (b) hybridizing a target nucleic acid sequence from
8 the first organism to the array; and

9 (c) based on an overall hybridization pattern of the
10 target to the array, identifying the genotype of the first
11 organism, and optionally identifying a phenotype of the first
12 organism.

1 2. The method of Claim 1, wherein the second
2 organism is *Mycobacterium tuberculosis*.

1 3. The method of Claim 2, wherein the reference
2 DNA or RNA sequences are selected from the group consisting of
3 16SrRNA, the *rpoB* gene, the *katG* gene, the *inhA* gene, the *gyrA*
4 gene, the 23SrRNA gene, the *rrs* gene, the *pnCA* gene, and the
5 *rpsL* gene.

1 4. The method of Claim 3, wherein the phenotype is
2 resistance to an antibiotic drug.

1 5. The method of Claim 4, wherein the drug is
2 selected from the group consisting of rifampacin, rifabutin,
3 isoniazid, streptomycin, pyrazinamide, ethambutol.

1 6. The method of Claim 1, wherein the overall
2 hybridization pattern is derived by comparing a hybridization
3 pattern of the target nucleic acid sequence to a hybridization
4 pattern of the reference sequence.

1 7. The method of Claim 6, wherein the comparing
2 identifies one or more positions at which a residue in the

3 target nucleic acid differs from a corresponding residue in
4 the reference sequence.

1 8. The method of Claim 7, wherein the comparing is
2 used to derive one or more sets of differences between the
3 target nucleic acid and the reference sequence, each set being
4 associated with a probability that the target belongs to a
5 particular species of the first organism.

1 9. The method of Claim 8, wherein the probability
2 associated with each set of differences is used to derive a
3 combined probability greater than a desired confidence level
4 that the target belongs to a particular species.

1 10. The method of Claim 8, wherein the comparing is
2 used to derive one or more sets of differences between the
3 target nucleic acid and the reference sequence, each set being
4 associated with a probability that the target possesses a
5 particular phenotype.

1 11. The method of Claim 10, wherein the probability
2 associated with each set of differences is used to derive a
3 combined probability greater than a desired confidence level
4 that the target possesses a particular phenotype.

1 12. The method of Claim 7, wherein the comparing
2 identifies one or more species-specific polymorphisms and
3 these species-specific polymorphisms are used to confirm the
4 identification.

1 13. The method of Claim 7, wherein the comparing
2 identifies one or more shared polymorphisms and these shared
3 polymorphisms are used to confirm the identification.

1 14. The method of Claim 6, wherein the
2 hybridization pattern of the target to a first region of the
3 array is used to derive a probability that the target belongs
4 to a particular species;

5 repeating this for other regions of the array until
6 the combination of probabilities derived from all the regions
7 indicating that the organism belongs to a particular species
8 exceeds a desired confidence level.

1 15. The method of Claim 14, wherein each region
2 corresponds to oligonucleotide probes which detect the
3 presence or absence of between three and fifteen contiguous
4 residues in the target nucleic acid.

1 16. The method of Claim 1, wherein the reference
2 DNA or RNA sequences are from a highly conserved gene.

1 17. The method of Claim 1, wherein the target
2 nucleic acid is amplified from a biological sample.

1 18. The method of Claim 17, wherein the target
2 nucleotide is fluorescently labelled.

1 19. The method of Claim 1, wherein the
2 oligonucleotides are from about 5 to 25 nucleotides in length.

1 20. The method of Claim 1, wherein the hybridizing
2 is performed in a fluid volume of 250 μ l or less.

1 21. The method of Claim 1, wherein the array has
2 between 100 and 1,000,000 probes.

1 22. The method of Claim 21, wherein the array has
2 approximately 2,800 probes.

1 23. The method of Claim 1, wherein the probes are
2 linked to the support via a spacer.

1 24. The method of Claim 1, wherein the overall
2 hybridization pattern is derived by:

3 (a) determining hybridization intensities of the
4 target nucleic acid sequence to each of a set of selected
5 probes; and

6 (b) comparing said hybridization intensities to
7 corresponding hybridization intensities of the reference
8 sequence to said set of selected probes.

1 25. The method of Claim 24, wherein the set of
2 selected probes interrogates a continuous segment of the
3 reference sequence.

1 26. The method of Claim 1, wherein the overall
2 hybridization pattern is derived by determining the maximum
3 hybridization intensity produced from a group of probes which
4 interrogate a common nucleotide position of the target
5 sequence, repeating this for other nucleotide positions in the
6 target, and plotting the determined maximum hybridization
7 intensities as a function of the corresponding nucleotide
8 position being interrogated to provide a target sequence plot
9 of hybridization intensity vs. nucleotide position.

1 27. The method of Claim 26, further comprising
2 repeating the steps of Claim 37 with the target sequence
3 replaced by the reference sequence, to derive a baseline plot
4 of the reference sequence and comparing the target plot to the
5 baseline plot.

1 28. The method of Claim 27, wherein the common
2 nucleotide positions form a continuous segment.

1 29. A method for identifying the genotype and/or
2 phenotype of an organism by comparing a target nucleic acid
3 sequence from a first organism coding for a gene (or its
4 complement) to a reference sequence coding for the same gene
5 (or its complement) from a second organism, said method
6 comprising:

7 (a) hybridizing a sample comprising the target
8 nucleic acid or a subsequence thereof to an array of

9 oligonucleotide probes immobilized on a solid support, the
10 array comprising:

11 a first probe set comprising a plurality of probes,
12 each probe comprising a segment of nucleotides exactly
13 complementary to a subsequence of the reference sequence, the
14 segment including at least one interrogation position
15 complementary to a corresponding nucleotide in the reference
16 sequence;

17 (b) determining which probes in the first probe set
18 bind to the target nucleic acid or subsequence thereof
19 relative to their binding to the reference sequence, such
20 relative binding indicating whether a nucleotide in the target
21 sequence is the same or different from the corresponding
22 nucleotide in the reference sequence;

23 (c) based on differences between the nucleotides of
24 the target sequence and the reference sequence identifying the
25 phenotype of the first organism;

26 (d) deriving one or more sets of differences
27 between the reference sequence and the first organism; and

28 (e) comparing the set of differences to a data base
29 comprising sets of differences correlated with speciation of
30 organisms to identify the genotype of the first organism.

1 30. The method of Claim 29, wherein the second
2 organism is *Mycobacterium tuberculosis*.

1 31. The method of Claim 29, wherein the gene is
2 selected from the group consisting of 16SrRNA, the *rpoB* gene,
3 the *katG* gene, the *inhA* gene, the *gyrA* gene, the 23SrRNA gene,
4 the *rrs* gene, the *pncA* gene, and the *rpsL* gene.

1 32. The method of Claim 29, wherein the phenotype
2 is resistance to an antibiotic drug.

1 33. The method of Claim 32, wherein the drug is
2 selected from the group consisting of rifampacin, rifabutin,
3 isoniazid, streptomycin, pyrazinamide, ethambutol.

1 34. The method of Claim 29, wherein the reference
2 DNA ord RNA sequences are from a highly conserved gene.

1 35. The method of Claim 29, wherein each set of
2 differences is associated with a probability that the target
3 belongs to a particular species of the first organism.

1 36. The method of Claim 35, wherein the probability
2 associated with each set of differences is used to derive a
3 combined probability greater than a desired confidence level
4 that the target belongs to a particular species.

1 37. The method of Claim 29, wherein the comparing
2 identifies one or more species-specific polymorphisms and
3 these species-specific polymorphisms are used to confirm the
4 identification.

1 38. The method of Claim 29, wherein the comparing
2 identifies one or more shared polymorphisms and these shared
3 polymorphisms are used to confirm the identification.

1 39. The method of Claim 29, wherein the target
2 nucleic acid is amplified from a biological sample.

1 40. The method of Claim 39, wherein the target
2 nucleic acid is fluorescently labelled.

1 41. The method of Claim 29, wherein the
2 oligonucleotides are from about 5 to 25 nucleotides in length.

1 42. The method of Claim 29, wherein the hybridizing
2 is performed in a fluid volume of 250 μ L or less.

1 43. The method of Claim 29, wherein the array has
2 between 100 and 1,000,000 probes.

1 44. The method of Claim 42, wherein the array has
2 approximately 2,800 probes.

3 45. The method of Claim 29, wherein the probes are
4 linked to the support via a spacer.

1 46. The method of Claim 29, wherein the array
2 further comprises a second, a third and a fourth probe sets
3 each comprising a corresponding probe for each probe in the
4 first probe set, the corresponding probes in the second, third
5 and fourth probe sets being identical in sequence to the
6 corresponding probe in the first probe set or a subsequence of
7 nucleotides thereof that includes the at least one
8 interrogation position, except that the at least one
9 interrogation position is occupied by a different nucleotide
10 in each of the four corresponding probes from the four probe
11 sets, and determining which probes, relative to one another,
12 in the four probe sets specifically bind to the target nucleic
13 acid or subsequence thereof, the relative specific binding of
14 the corresponding probes in the four probe sets indicating
15 whether a nucleotide in the target sequence is the same or
16 different from the corresponding nucleotide in the reference
17 sequence.

1 47. The method of Claim 46, wherein the array
2 further comprises a fifth probe set comprising a corresponding
3 probe for each probe in the first probe set, the corresponding
4 probe from the fifth probe set being identical to a sequence
5 comprising the corresponding probe from the first probe set or
6 a subsequence of nucleotides thereof that includes the at
7 least one interrogation position, except that the at least one
8 interrogation position is deleted in the corresponding probe
9 from the fifth probe set.

1 48. The method of Claim 46, wherein the array
2 further comprises a sixth probe set comprising a corresponding
3 probe for each probe in the first probe set, the corresponding
4 probe from the sixth probe set being identical to a sequence
5 comprising the corresponding probe from the first probe set or
6 a subsequence of nucleotides thereof that includes the at
7 least one interrogation position, except that an additional

8 nucleotide is inserted adjacent to the at least one
9 interrogation position in the corresponding probe from the
10 first probe set.

1 49. The method of Claim 46, wherein the first probe
2 set has at least three interrogation positions respectively
3 corresponding to each of three contiguous nucleotides in a
4 reference sequence.

1 50. The method of Claim 46, wherein the first probe
2 set has at least 50 interrogation positions respectively
3 corresponding to each of 50 contiguous nucleotides in a
4 reference sequence.

1 51. The method of Claim 46, wherein the segment in
2 each probe of the first probe set that is exactly
3 complementary to the subsequence of the reference sequence is
4 9-21 nucleotides.

1 52. A method for identifying the genotype and/or
2 phenotype of an organism by comparing a target nucleic acid
3 sequence from a first organism coding for a gene (or its
4 complement) to a reference sequence coding for the same gene
5 (or its complement) from a second organism, said method
6 comprising:

7 (a) hybridizing a sample comprising the target
8 nucleic acid or a subsequence thereof to an array of
9 oligonucleotide probes immobilized on a solid support, the
10 array comprising:

11 a first probe set comprising a plurality of probes,
12 each probe comprising a segment of nucleotides exactly
13 complementary to a subsequence of the reference sequence, the
14 segment including at least one interrogation position
15 complementary to a corresponding nucleotide in the reference
16 sequence, wherein each interrogation position corresponds to a
17 nucleotide position in the reference or target sequence;

18 (b) determining a hybridization intensity from each
19 probe;

(c) plotting the hybridization intensities versus the nucleotide position corresponding to the probe from which the hybridization intensity was determined to derive a target plot of hybridization intensity;

(d) repeating steps (a) - (c) with the target sequence replaced by the reference sequence, to derive a baseline plot of the reference sequence; and

(e) comparing the target plot to the baseline plot to identify the genotype and/or phenotype of the organism.

53. The method of Claim 52, wherein the second organism is *Mycobacterium tuberculosis*.

54. The method of Claim 53, wherein the gene is selected from the group consisting of 16SrRNA, the *rpoB* gene, the *katG* gene, the *inhA* gene, the *gyrA* gene, the 23SnRNA gene, the *rrs* gene, the *pnCA* gene, and the *rpsL* gene.

55. The method of Claim 54, wherein the phenotype is resistance to an antibiotic drug.

56. The method of Claim 55, wherein the drug is selected from the group consisting of rifampacin, rifabutin, isoniazid, streptomycin, pyrazinamide, ethambutol.

57. The method of Claim 52, wherein the reference DNA or RNA sequences are from a highly conserved gene.

58. The method of Claim 52, wherein the array further comprises a second, a third and a fourth probe sets each comprising a corresponding probe for each probe in the first probe set, the corresponding probes in the second, third and fourth probe sets being identical in sequence to the corresponding probe in the first probe set or a subsequence of nucleotides thereof that includes the at least one interrogation position, except that the at least one interrogation position is occupied by a different nucleotide in each of the four corresponding probes from the four probe

11 sets, and the hybridization intensity determined in (b) is the
12 maximum hybridization intensity from each of the corresponding
13 probes in the four probe sets.

1 59. An array of oligonucleotide probes immobilized
2 on a solid support, the array comprising:

3 a first probe set comprising a plurality of probes,
4 each probe comprising a segment of nucleotides exactly
5 complementary to a subsequence of a reference sequence, the
6 segment including at least one interrogation position
7 complementary to a corresponding nucleotide in the reference
8 sequence;

9 wherein the reference sequence is a gene from
10 *Mycobacterium tuberculosis*.

1 60. The array of Claim 59, wherein the gene is
2 selected from the group consisting of 16SrRNA, the *rpoB* gene,
3 the *katG* gene, the *inhA* gene, the *gyrA* gene, the 23SrRNA gene,
4 the *rrs* gene, the *pncA* gene, and the *rpsL* gene.

1 61. The array of Claim 60, further comprising:
2 a second, a third and a fourth probe sets each
3 comprising a corresponding probe for each probe in the first
4 probe set, the corresponding probes in the second, third and
5 fourth probe sets being identical in sequence to the
6 corresponding probe in the first probe set or a subsequence of
7 nucleotides thereof that includes the at least one
8 interrogation position, except that the at least one
9 interrogation position is occupied by a different nucleotide
10 in each of the four corresponding probes from the four probe
11 sets.

1 62. A method of identifying the presence of a
2 nucleic acid polymorphism in a patient sample, comprising the
3 steps of:

4 (a) determining the difference between the
5 hybridization intensities of a nucleic acid sequence from the
6 patient sample and a corresponding nucleic acid sequence from

7 a wild type sample to an array of reference nucleic acid
8 probes;

9 (b) deriving ratios of the difference in (a) to the
10 hybridization intensity of the wild type sample for each base
11 position corresponding to each reference nucleic acid probe;
12 and

13 (c) identifying the presence of a polymorphism at a
14 base position corresponding to a reference probe if the ratio
15 in (b) for the base position corresponding to the reference
16 probe is greater than or equal to an assigned value.

1 63. The method of claim 62, wherein the
2 nucleic acid sequence is selected from the group consisting of
3 mitochondrial DNA, p53, MSH, MLH1, or BRCA-1.

1 64. The method of claim 62, wherein the
2 nucleic acid sequence comprises an HIV gene.

1 65. The method of claim 62, wherein the
2 nucleic acid sequence comprises a gene associated with a
3 heritable disease.

1 66. The method of claim 65, wherein the
2 heritable disease is cystic fibrosis.

1 67. A computer program product that identifies the
2 presence of a nucleic acid polymorphism in a patient sample,
3 comprising:

4 computer code that determines the difference between
5 the hybridization intensities of a nucleic acid sequence from
6 the patient sample and a corresponding nucleic acid sequence
7 from a wild type sample to an array of reference nucleic acid
8 probes;

9 computer code that derives ratios of the difference
10 to the hybridization intensity of the wild type sample for
11 each base position corresponding to each reference nucleic
12 acid probe;

13 computer code that identifies the presence of a
14 polymorphism at a base position corresponding to a reference
15 probe if the ratio for the base position corresponding to the
16 reference probe is greater than or equal to an assigned value;
17 and

18 a computer readable medium that stores the computer
19 codes.

1 68. In a computer system, a method of assigning an
2 organism to a group, comprising the steps of:

3 inputting groups of a plurality of known nucleic
4 acid sequences, the plurality of known nucleic acid sequences
5 being from known organisms;

6 inputting hybridization patterns for the plurality
7 of known nucleic acid sequences, each hybridization pattern
8 indicating hybridization of subsequences of the known nucleic
9 acid sequence to subsequences of a reference nucleic acid
10 sequence;

11 inputting a hybridization pattern for a sample
12 nucleic acid sequence from the organism indicating
13 hybridization of subsequences of the sample nucleic acid
14 sequence to subsequences of the reference nucleic acid
15 sequence;

16 comparing the hybridization pattern for the sample
17 nucleic acid sequence to the hybridization patterns for the
18 plurality of known nucleic acid sequences; and

19 assigning a particular group to which the organism
20 belongs according to the group of at least one of the known
21 nucleic acid sequences that has a hybridization pattern that
22 most closely matches the hybridization pattern of the sample
23 nucleic acid sequence at specific locations.

1 69. The method of claim 68, wherein the group is
2 selected from the group consisting of species, subspecies,
3 genotype, and phenotype.

1 70. The method of claim 68, wherein the group to
2 which a sample nucleic acid sequence is assigned is determined

3 without requiring knowledge of the actual nucleotide sequence
4 of the sample nucleic acid sequence.

1 71. The method of claim 68, further comprising the
2 step of normalizing hybridization intensities of the
3 hybridization patterns of the sample and known nucleic acid
4 sequences using linear regression.

1 72. The method of claim 71, wherein the comparing
2 step include utilizing a regression coefficient from the
3 linear regression for comparison.

1 73. The method of claim 68, further comprising the
2 step of generating a database of the hybridization patterns
3 for the plurality of known nucleic acid sequences.

1 74. The method of claim 68, wherein the reference
2 nucleic acid sequence is from *Mycobacterium tuberculosis*.

1 75. The method of claim 68, wherein the locations
2 include locations of species-specific polymorphisms.

1 76. The method of claim 68, wherein the locations
2 include locations of shared polymorphisms between or among
3 multiple species.

1 77. The method of claim 68, further comprising the
2 step of calculating a probability that the sample nucleic acid
3 sequence belongs to the particular group.

1 78. The method of claim 68, wherein the group is a
2 species of *Mycobacterium*.

1 79. The method of claim 68, wherein the known and
2 sample nucleic acid sequences include a highly conserved gene.

1 80. The method of claim 68, wherein the known and
2 sample nucleic acid sequences include a gene selected from the

3 group consisting of 16SrRNA, the *rpoB* gene, the *katG* gene, the
4 *inhA* gene, the *gyrA* gene, the 23SrRNA gene, the *rrs* gene, the
5 *pncA* gene, and the *rpsL* gene.

1 81. A computer program product that assigns an
2 organism to a group, comprising:

3 computer code that receives as input groups of a
4 plurality of known nucleic acid sequences, the plurality of
5 known nucleic acid sequences being from known organisms;

6 computer code that receives as input hybridization
7 patterns for the plurality of known nucleic acid sequences,
8 each hybridization pattern indicating hybridization of
9 subsequences of the known nucleic acid sequence to
10 subsequences of a reference nucleic acid sequence;

11 computer code that receives as input a hybridization
12 pattern for a sample nucleic acid sequence from the organism
13 indicating hybridization of subsequences of the sample nucleic
14 acid sequence to subsequences of the reference nucleic acid
15 sequence;

16 computer code that compares the hybridization
17 pattern for the sample nucleic acid sequence to the
18 hybridization patterns for the plurality of known nucleic acid
19 sequences;

20 computer code that assigns a particular group to
21 which the organism belongs according to the groups of at least
22 one of the known nucleic acid sequences that has a
23 hybridization pattern that most closely matches the
24 hybridization pattern of the sample nucleic acid sequence at
25 specific locations; and

26 a computer readable medium that stores the computer
27 codes.

1 82. In a computer system, a method of assigning
2 groups to which organisms belong utilizing a generic probe
3 array, comprising the steps of:

4 inputting hybridization intensities for a plurality
5 of isolates, the hybridization intensities indicating

6 hybridization affinity between the isolate and the generic
7 probe array;

8 selecting hybridization intensities that have the
9 most variance across the plurality of isolates; and

10 assigning each of the plurality of isolates to a
11 group according to the selected hybridization intensities.

2 83. The method of claim 82, wherein the group is
3 selected from the group consisting of species, subspecies,
genotype, and phenotype.

1 84. The method of claim 82, wherein the assigning
2 step comprises the step of clustering the plurality of
3 isolates into groups according to the selected hybridization
4 intensities.

1 85. The method of claim 84, wherein the clustering
2 step is selected from the group consisting of Principal
3 Components analysis and Variable Clustering analysis.

1 86. The method of claim 82, further comprising the
2 step of standardizing the hybridization intensities among the
3 plurality of isolates.

2 87. The method of claim 86, wherein the
3 standardizing step comprises the step of adjusting the
4 hybridization intensities of each isolate so that there is a
common mean and variance across the plurality of isolates.

1 88. The method of claim 82, wherein the generic
2 probe array includes all nucleic acid probes of a specific
3 length.

1 89. A computer program product that assigns groups
2 to which organisms belong utilizing a generic probe array,
3 comprising the steps of:

4 computer code that receives as input hybridization
5 intensities for a plurality of isolates, the hybridization

6 intensities indicating hybridization affinity between the
7 isolate and the generic probe array;
8 computer code that selects hybridization intensities
9 that have the most variance across the plurality of isolates;
10 computer code that assigns a group to each of the
11 plurality of isolates according to the selected hybridization
12 intensities; and
13 a computer readable medium that stores the computer
14 codes.

1 / 3 6

CORRESPONDING
NUCLEOTIDE

ACTGTTAGCTAATTGG — REF. SEQ.
 CAATCGA — PROBE FROM FIRST PROBE SET
 CAATCGA — CORRESPONDING PROBES
 CAATCGA — FROM SECOND, THIRD AND
 CAATCGA — FOURTH PROBE SETS

INTERROGATION
POSITION

FIG. 1

ACTGTTAGCTAATTGG — REF. SEQ.
 GGGCAATCGAGGGGGG — PROBE FROM FIRST PROBE SET

LEADING SEGMENT OF TRAILING
SEQUENCE COMPLEMENTARITY SEQUENCE

FIG. 2

2 / 3 6

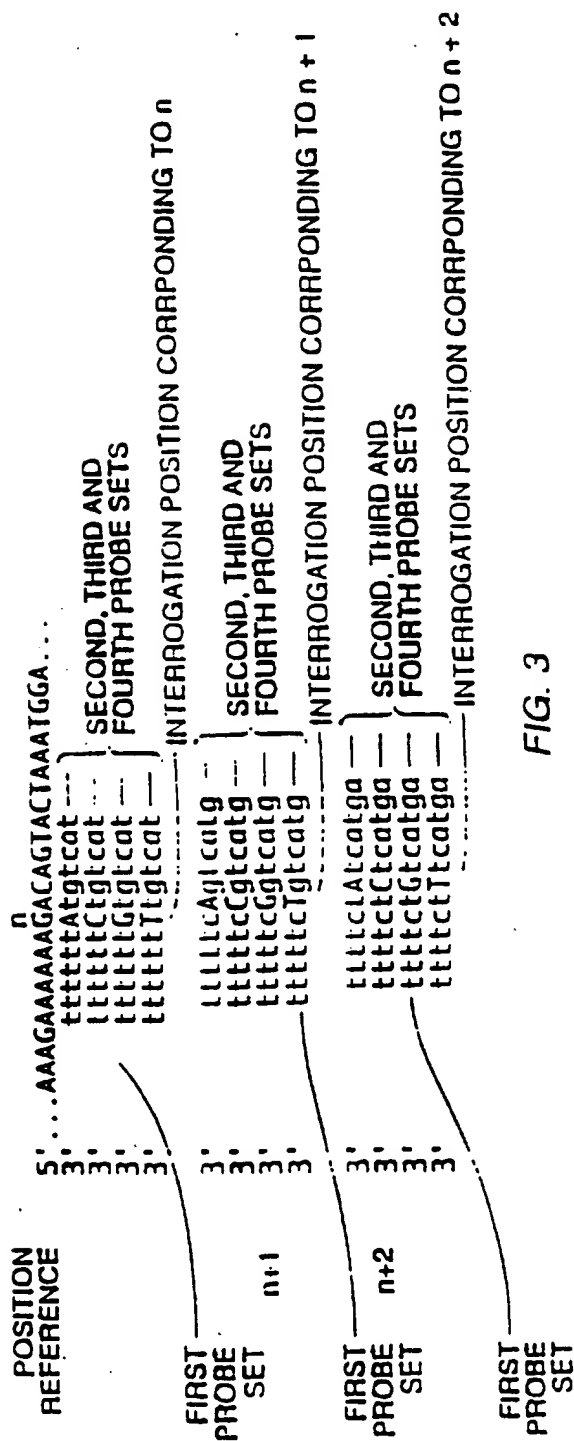


FIG. 3

3 / 3 6

$\begin{matrix} n_1 n_2 n_3 n_4 n_5 \\ \text{ACTGTTAGCTAATTGG} \end{matrix}$ REF. SEQ.

A-LANE	TGAC	GATA	ACCA	CAAT	AAAG
C-LANE	TGCC	GACA	ACCA	CACT	AAAG
G-LANE	TGCC	GACA	ACCA	CACT	AAAG
T-LANE	TGTC	GATA	ACCA	CAAT	AAAG
	1	2	3	4	5

WT. LANE TGAC GACA ACCA CAAT AAAG

FIG. 4

$\begin{matrix} n \\ \text{ACTGTTAGCTAATTGG} \end{matrix}$ REF. SEQ.

CAATCGA	PROBE FROM FIRST SET
CAATCGAT	DELETION PROBE
CAATACGA	INSERTION PROBES
CAATCCGA	
CAATGCCA	
CAATTCGA	

FIG. 5

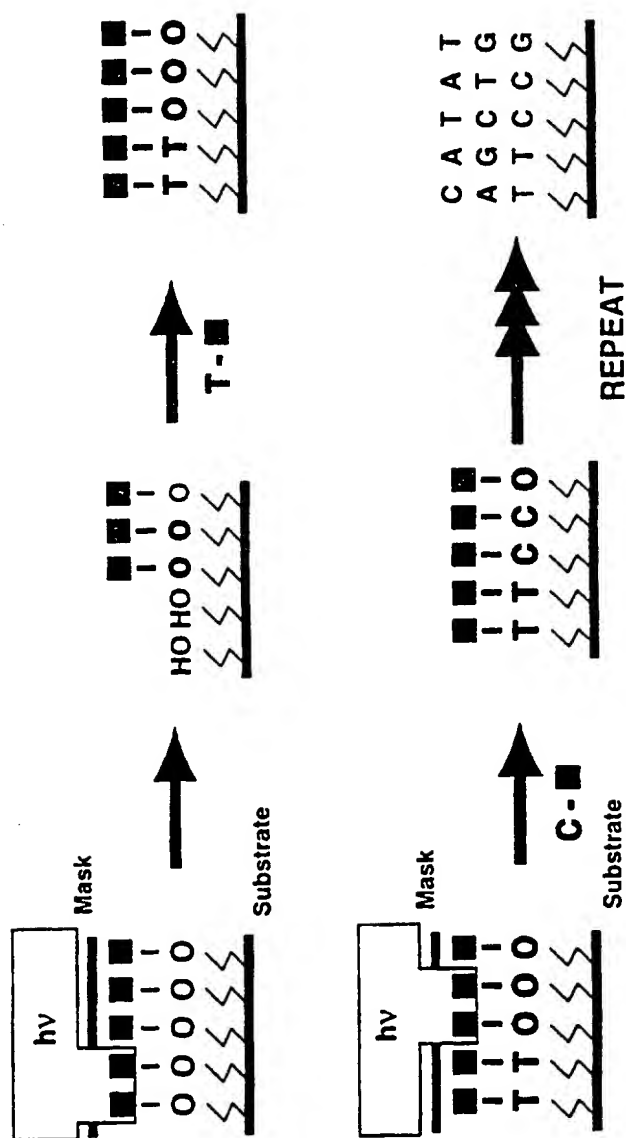


Figure 6

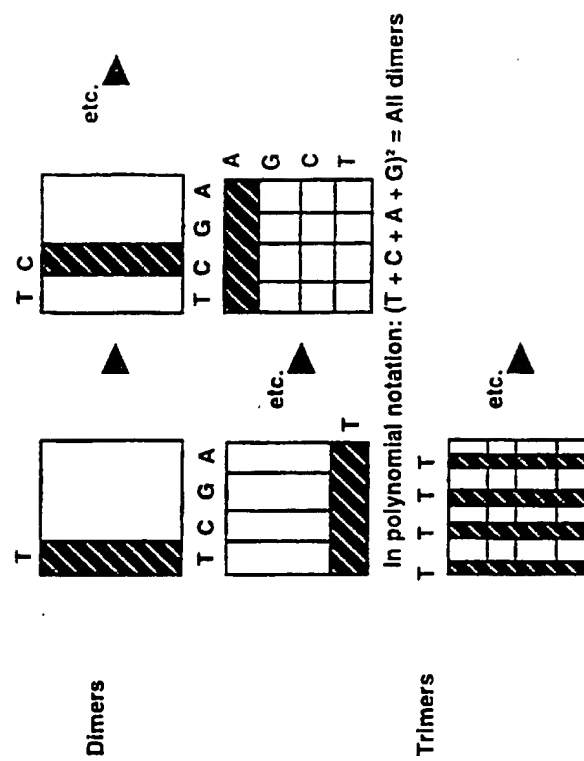


Figure 7

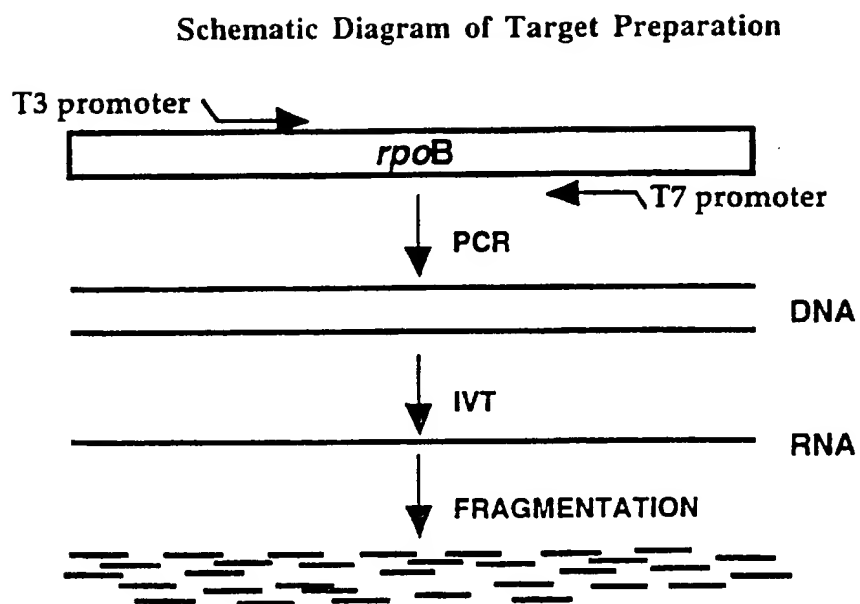


Figure 8

7 / 3 6

Tiling Strategy for Sequence Determination

Non Coding Sequences							Coding Sequences													
-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14
							A													
							C													
							G													
							T													
								A												
								C												
								G												
								T												
									A											
									C											
									G											
									T											
										A										
										C										
										G										
										T										
							1	2	3	4										

Figure 9

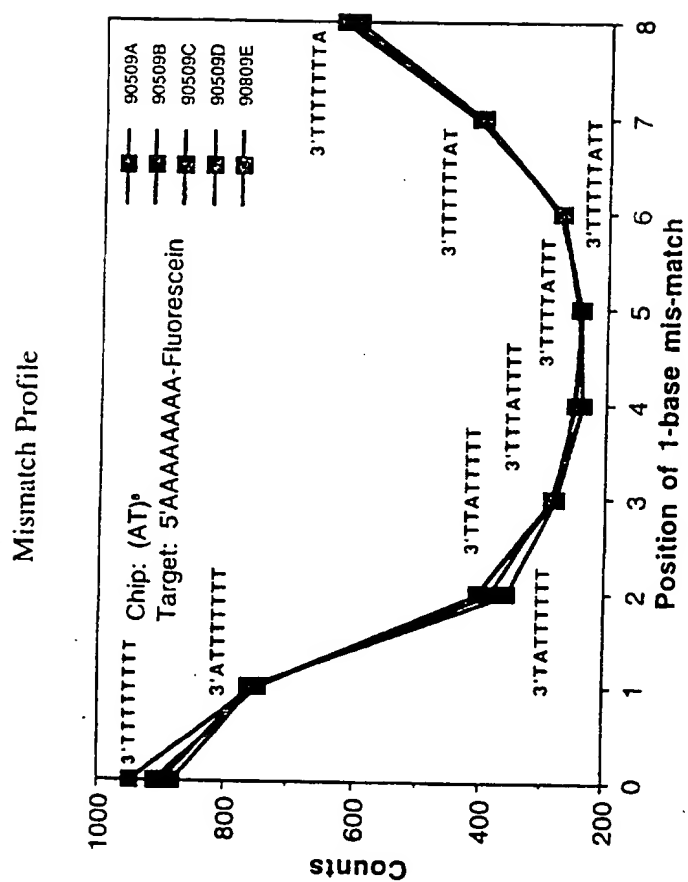


Figure 10

1 0 / 3 6

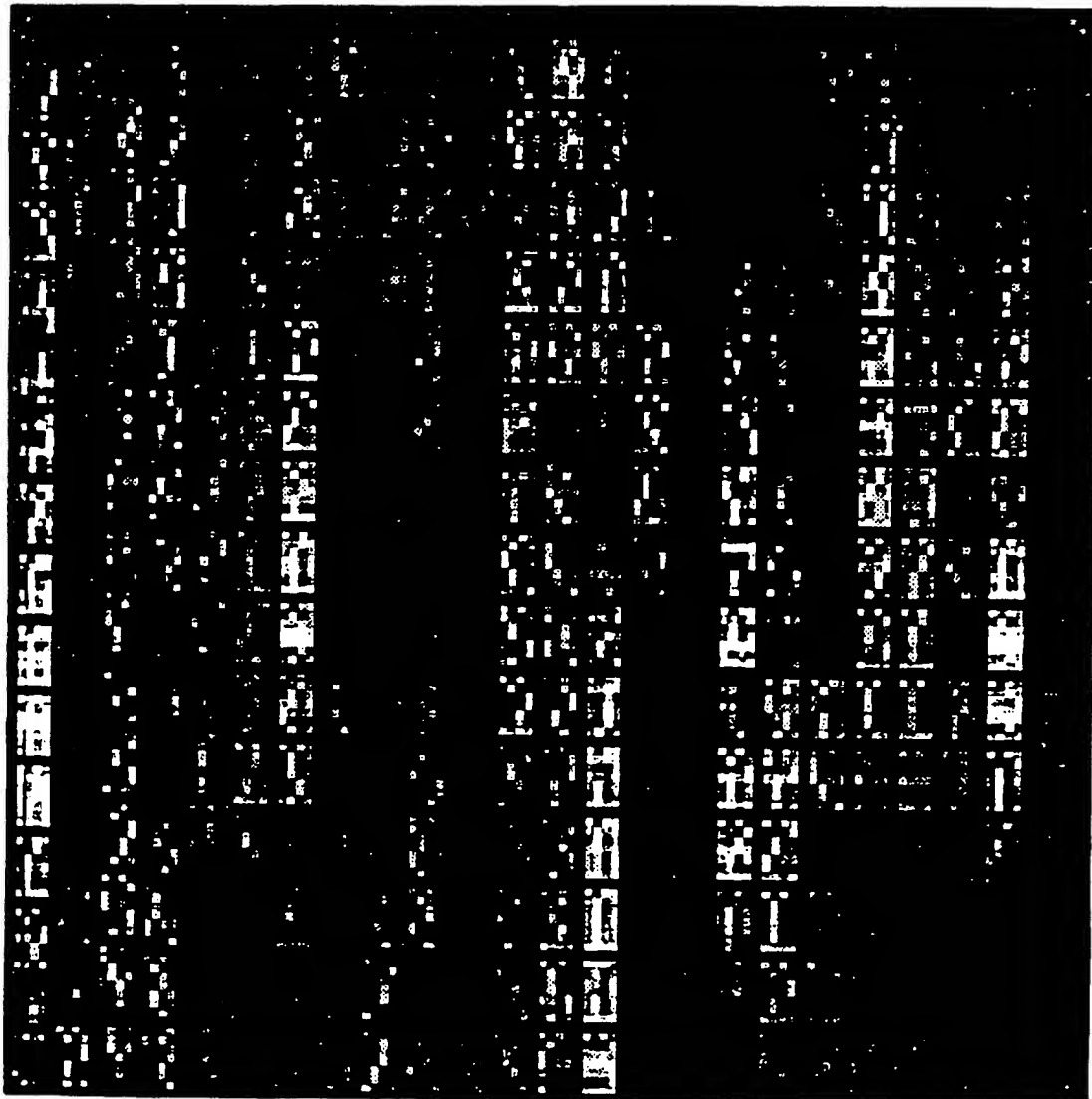


Figure 12

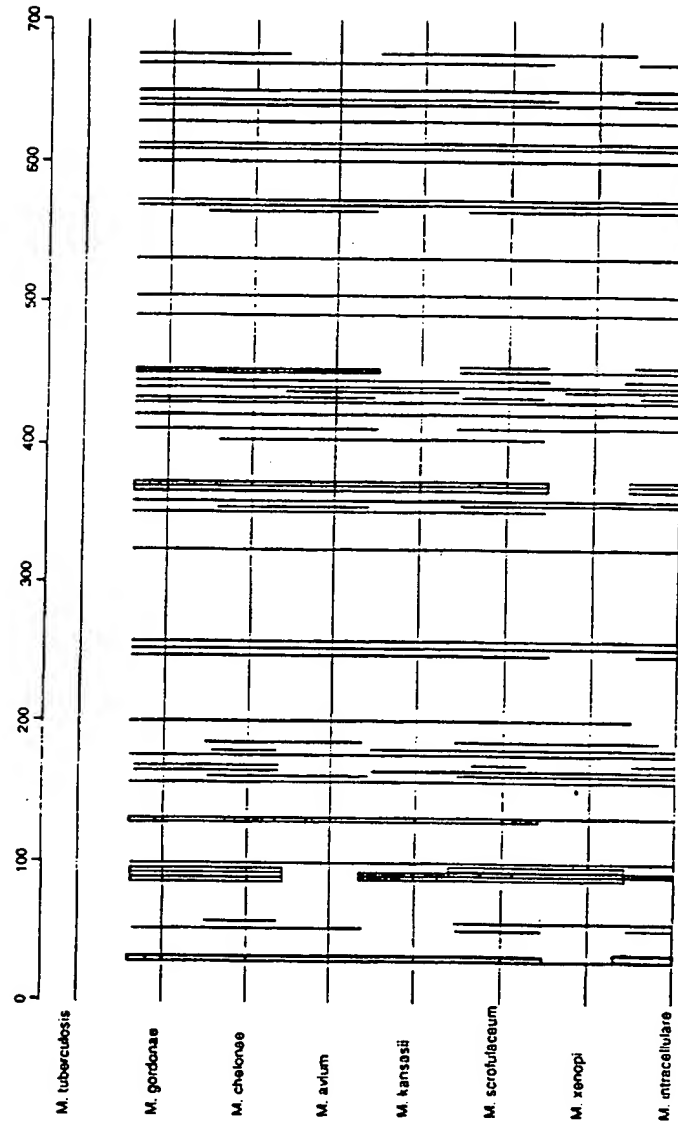


Figure 13

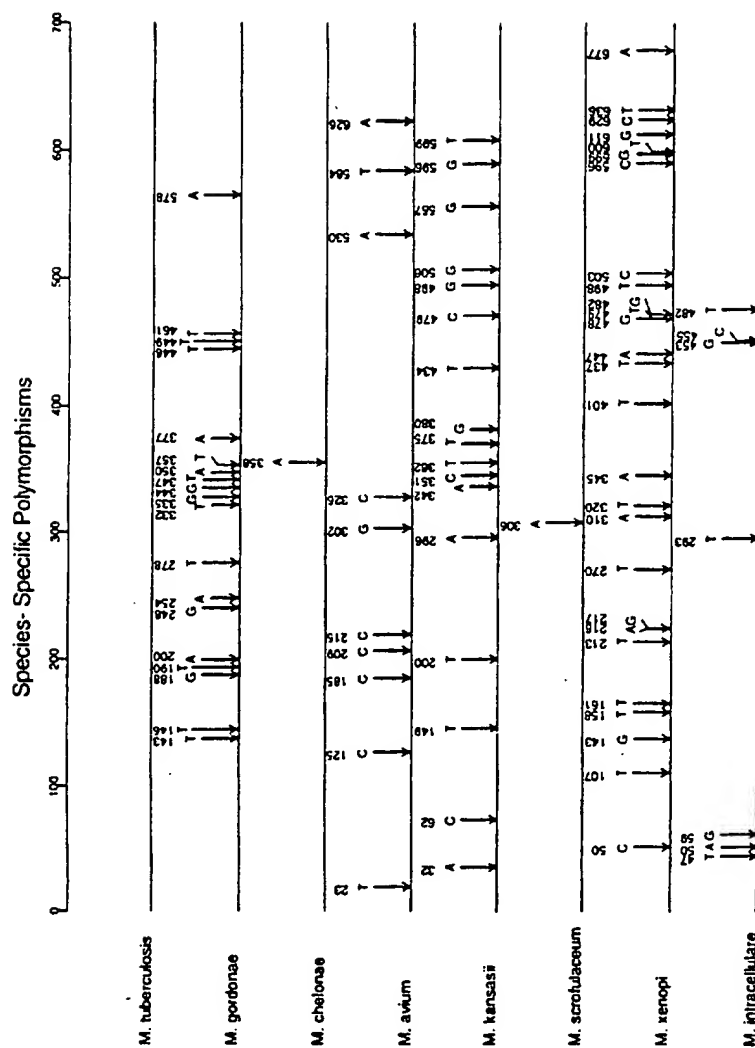


Figure 14

13 / 36

Fingerprints of Mycobacterium Species



M. chelonae



M. gordonae



M. xenopi



M. scrofulaceum



M. avium



M. kansasii



M. avium/intracellulare

M. tuberculosis

Figure 15A

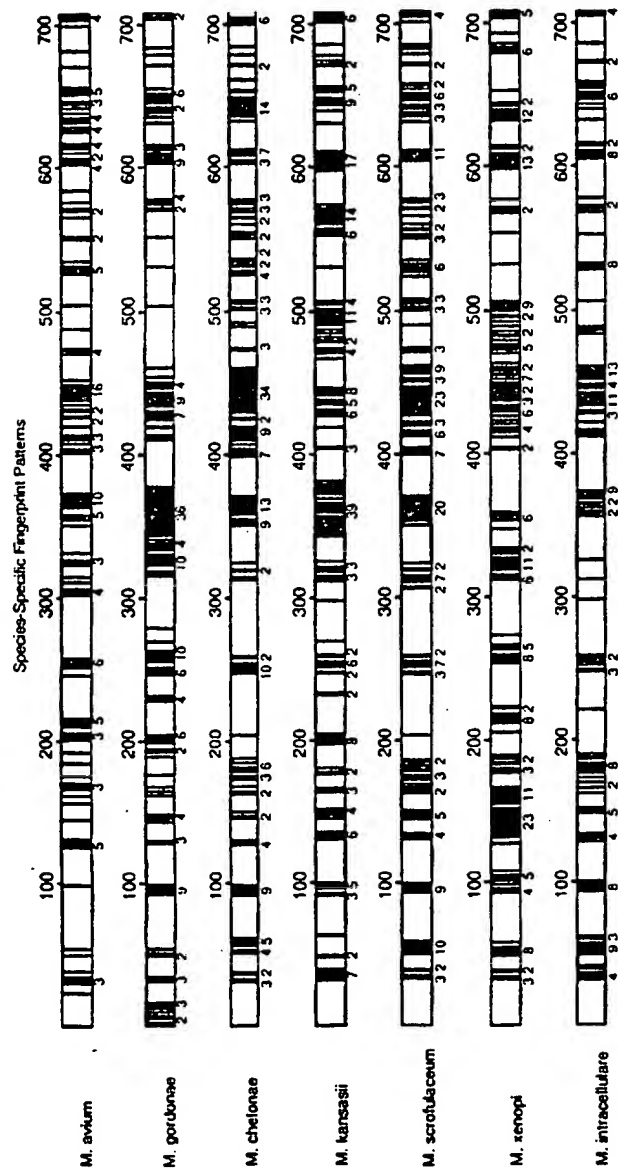


Figure 15B

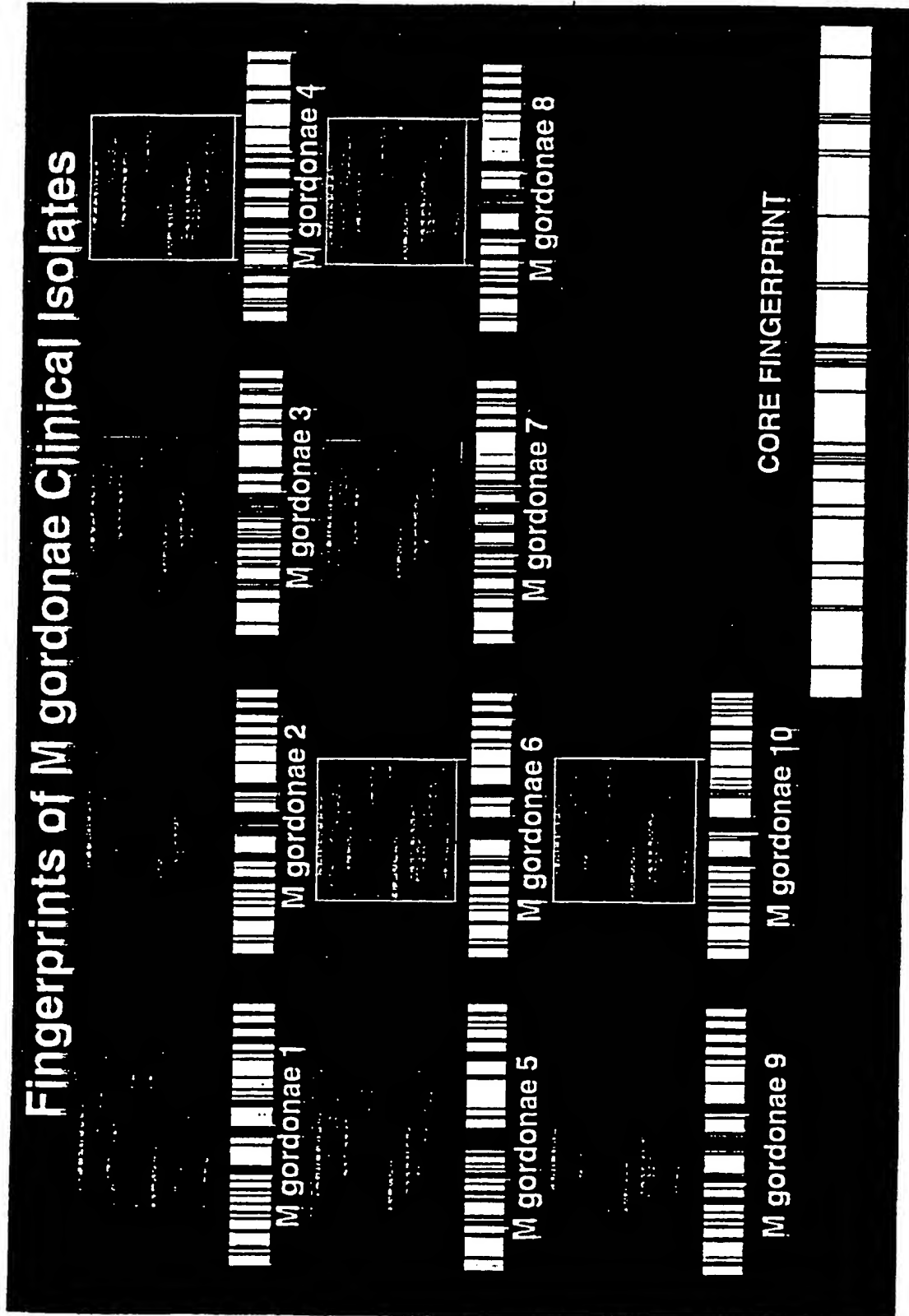


Figure 16

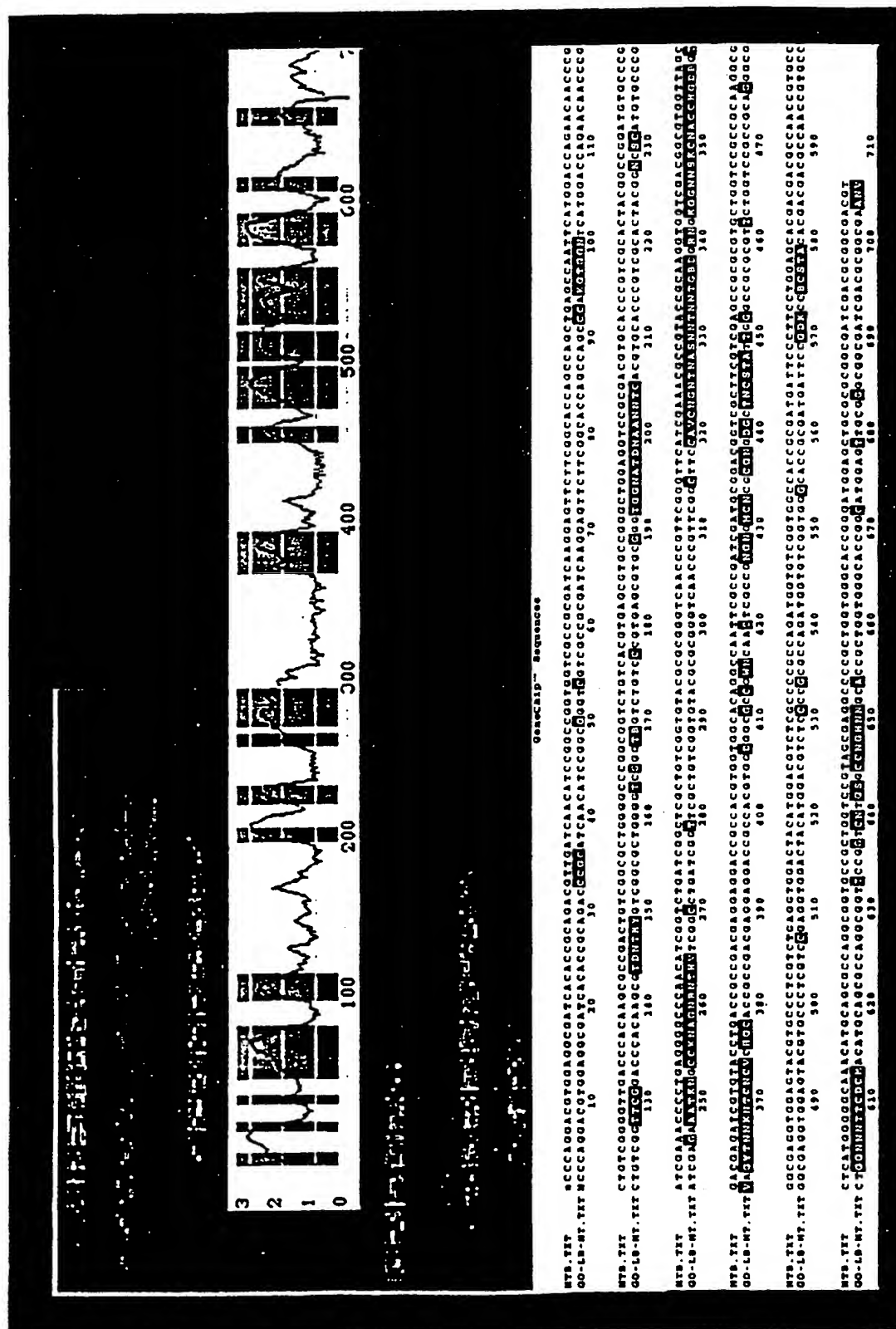


Figure 17

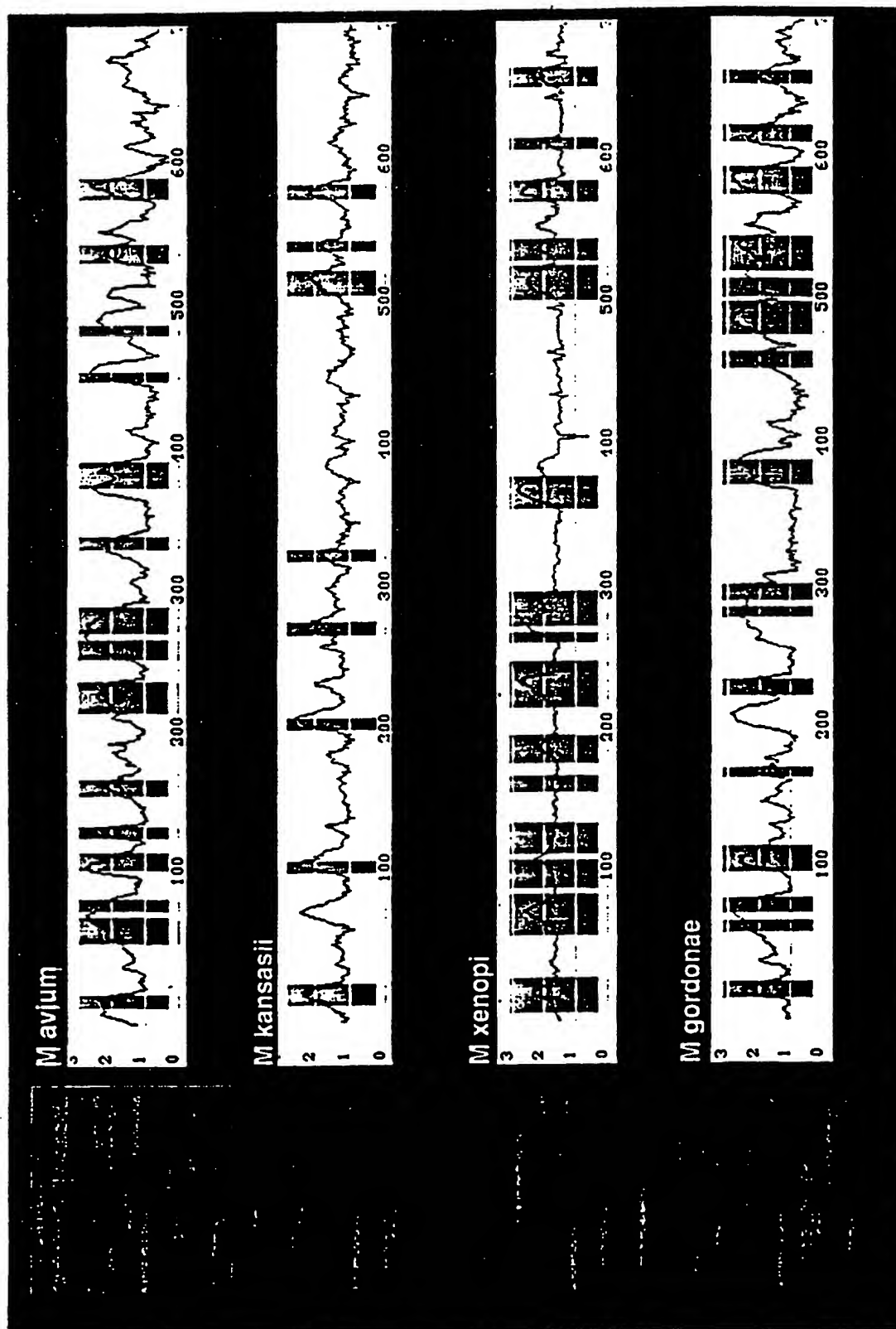


Figure 18

18 / 36

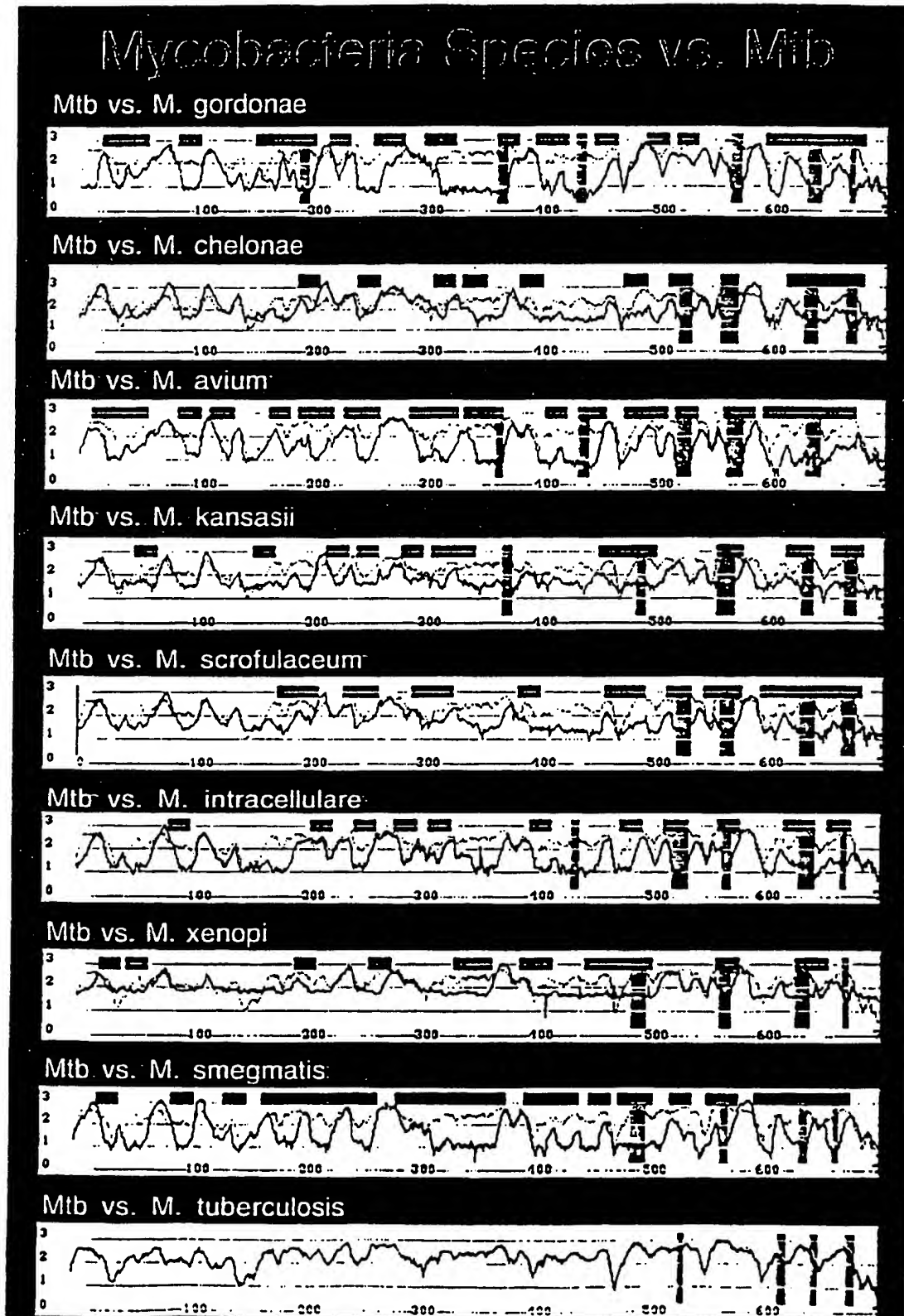


Figure 19

No Match

M avium

Patient A

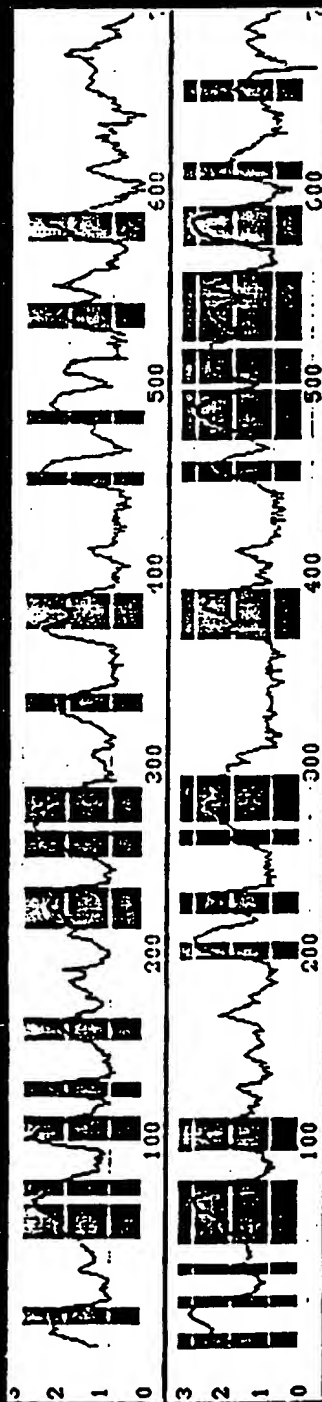
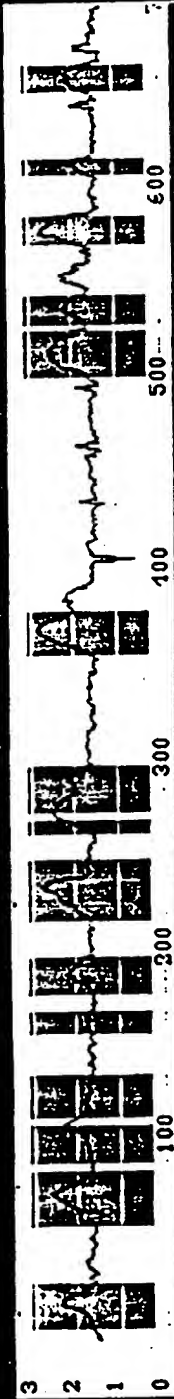


Figure 20A

20 / 36

No Match

M xenopi



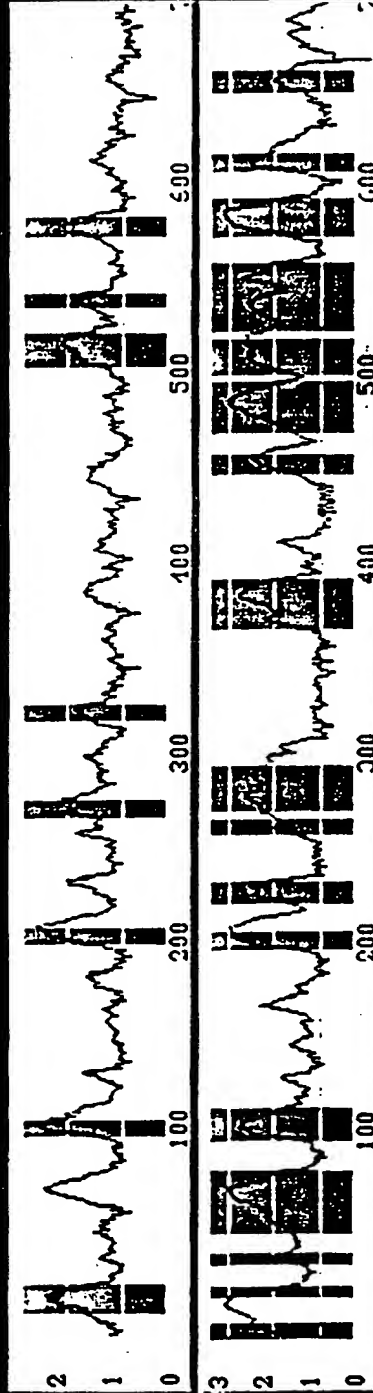
Patient A



Figure 20B

No Match

M kansasii

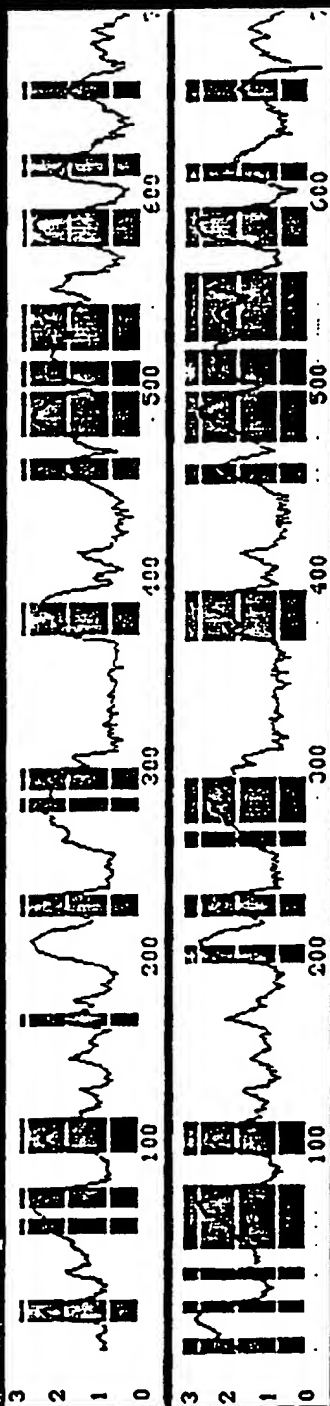


Patient A

Figure 20C

Match

M. gordonae



Patient A

Figure 20D

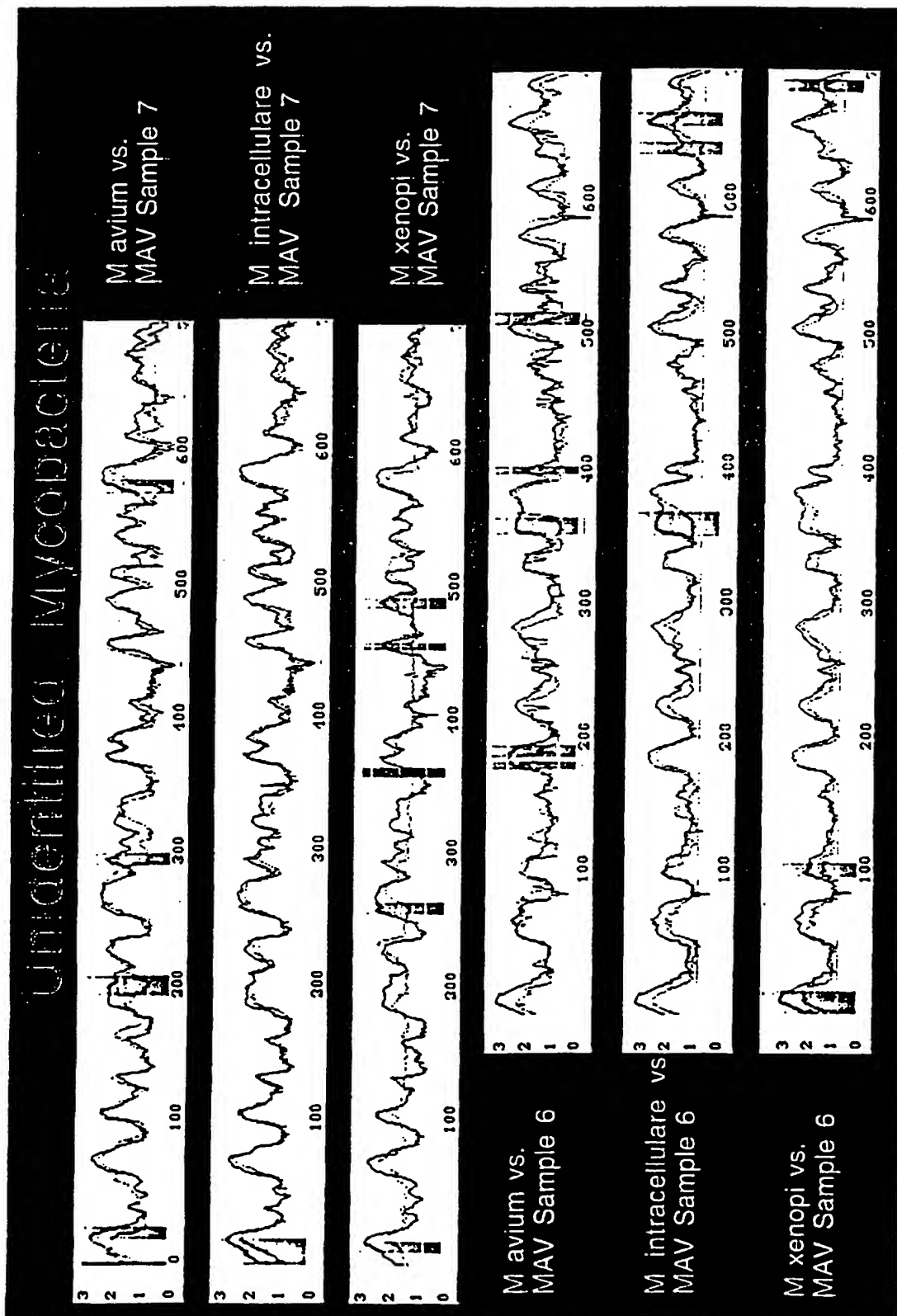


Figure 21

24 / 36

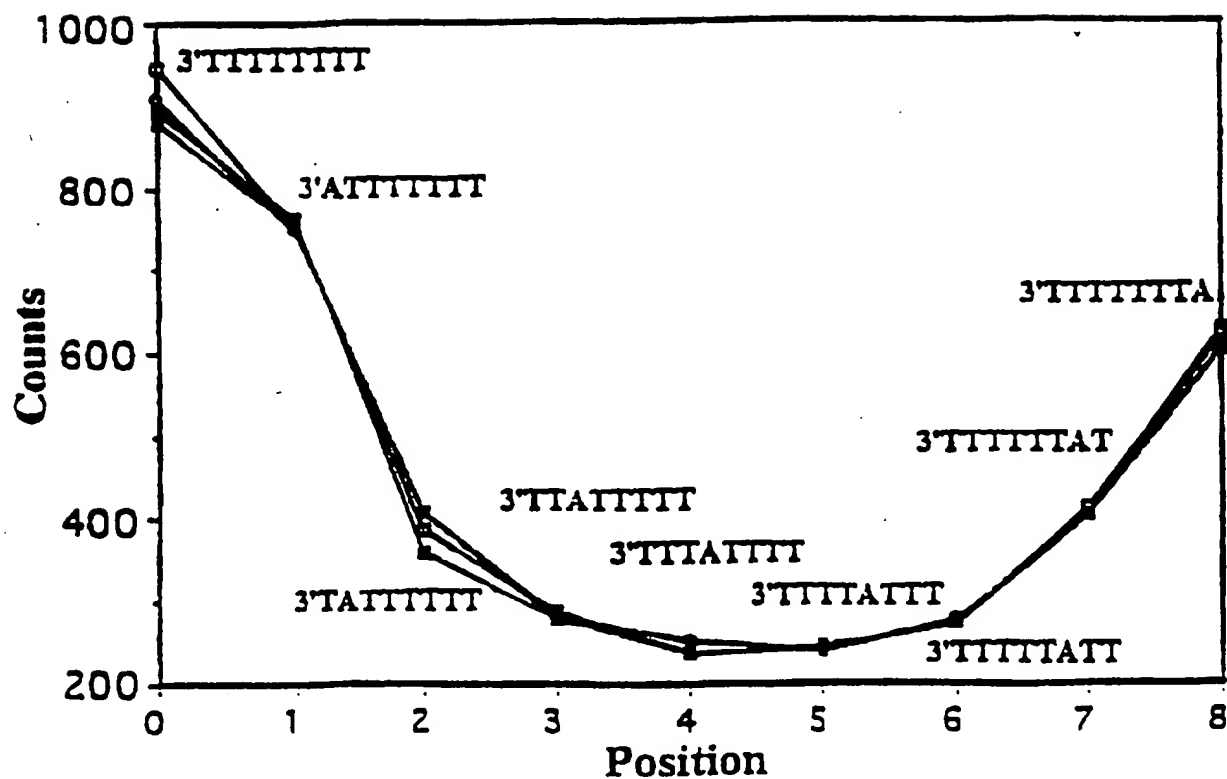
M. gordonae isolates*M. gordonae* (ATCC) vs. *M. gordonae* LB*M. gordonae* (ATCC) vs. *M. gordonae* WN*M. gordonae* (ATCC) vs. *M. gordonae* RM*M. gordonae* (ATCC) vs. *M. gordonae* LZ

Figure 22

Fig. 23



Fig. 24



26 / 36

FIG. 25

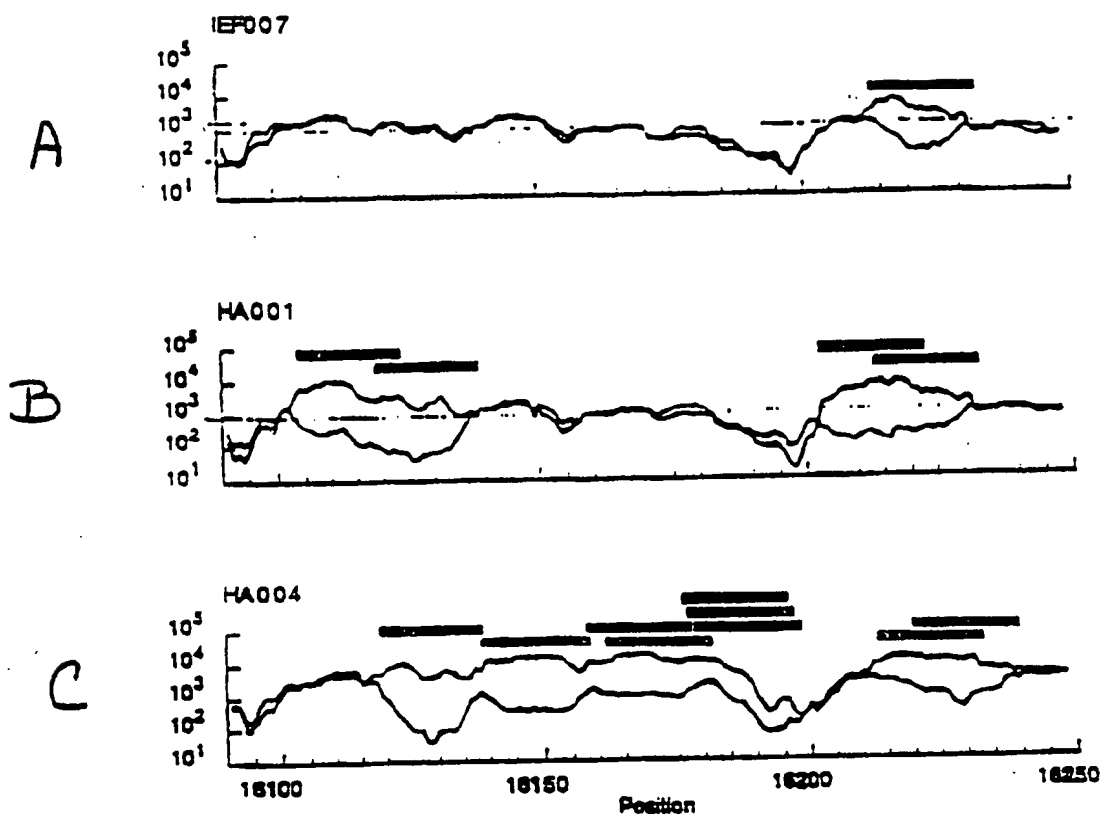
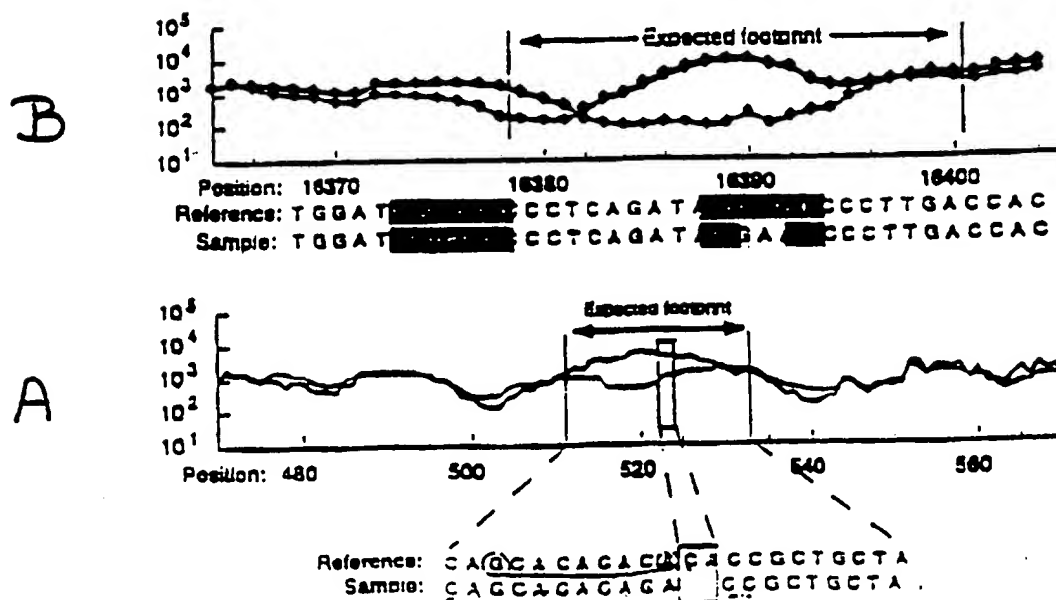


FIG. 26



27 / 36

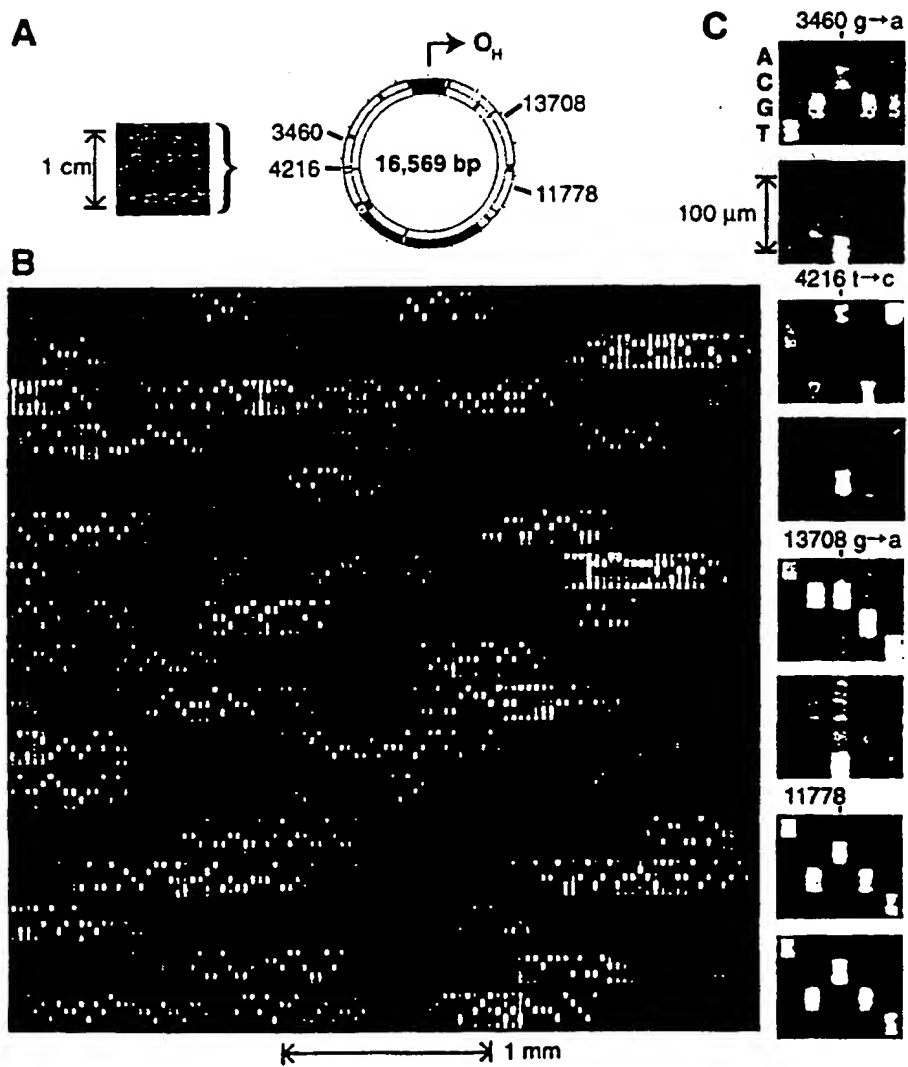
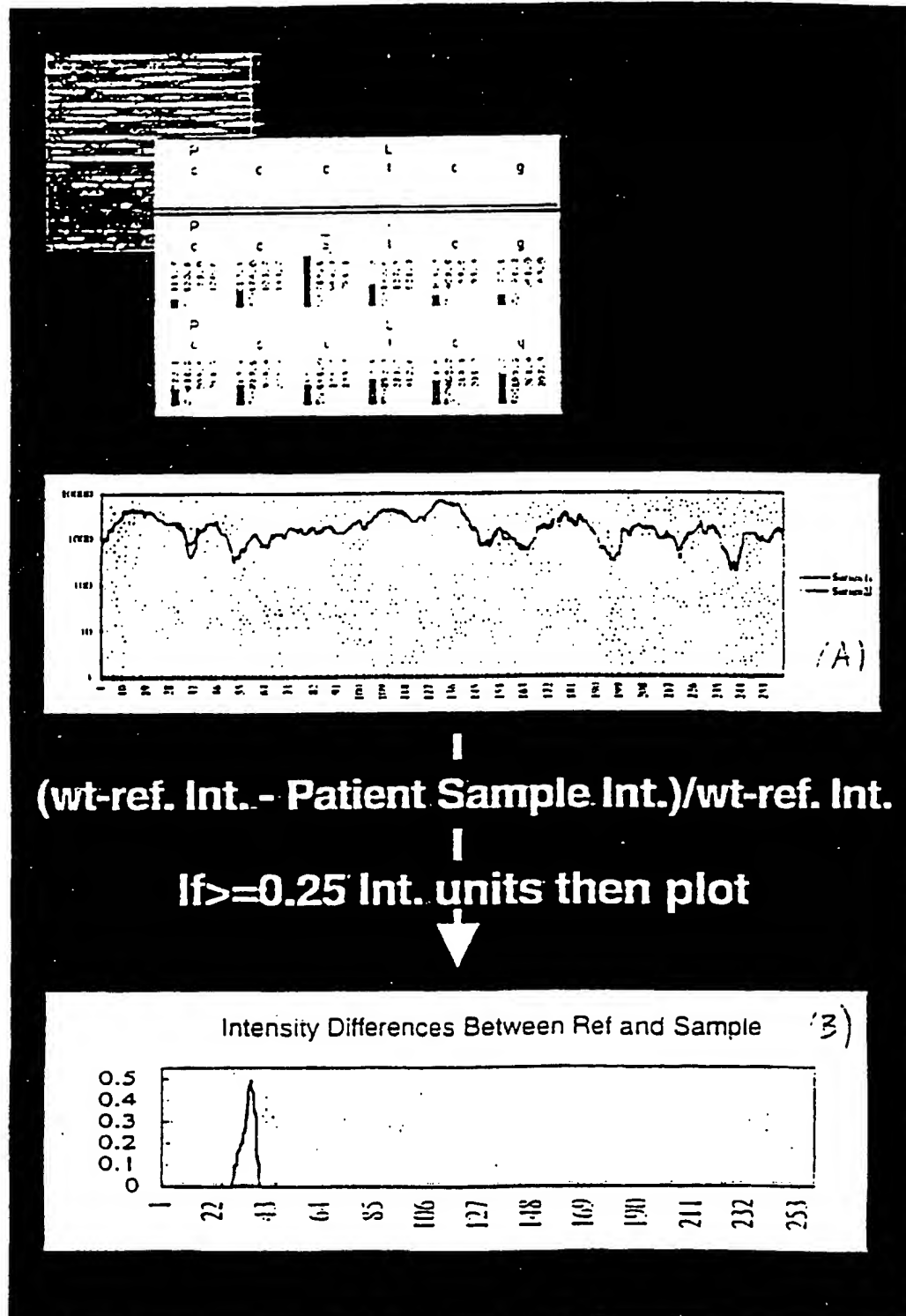


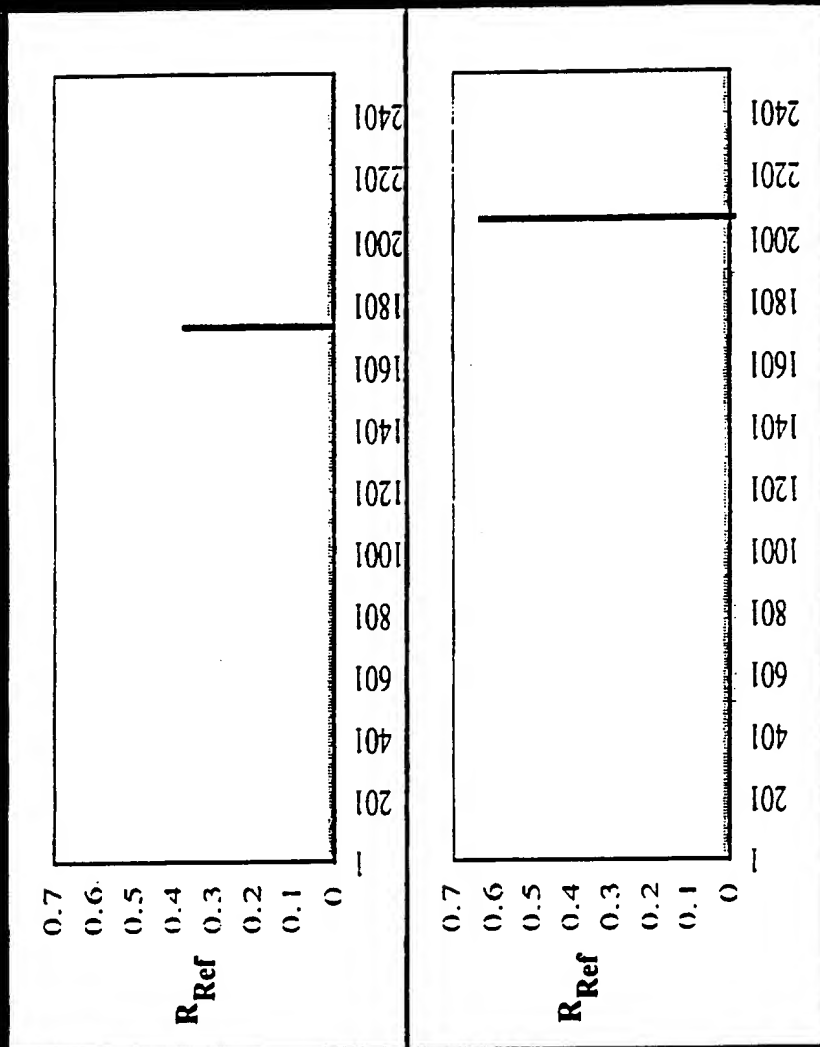
Figure 27

28 / 36

FIG. 28



Detection of Polynucleotide Mutations in DNA



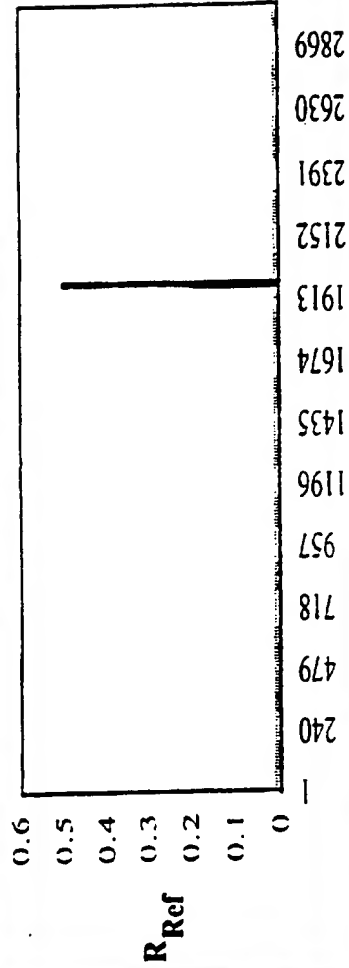
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----

Figure 29

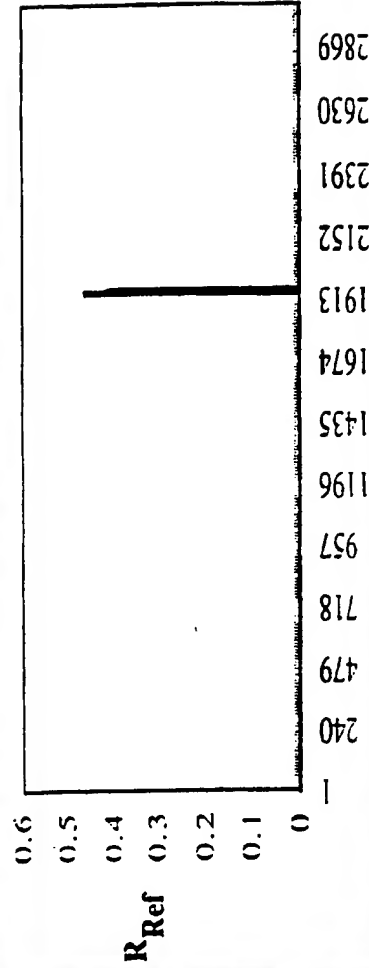
FIG-30

Detection of Heterozygous Mutations in P13H2

C -> T substitution
in Exon 12



3 bp deletion in
Exon 12

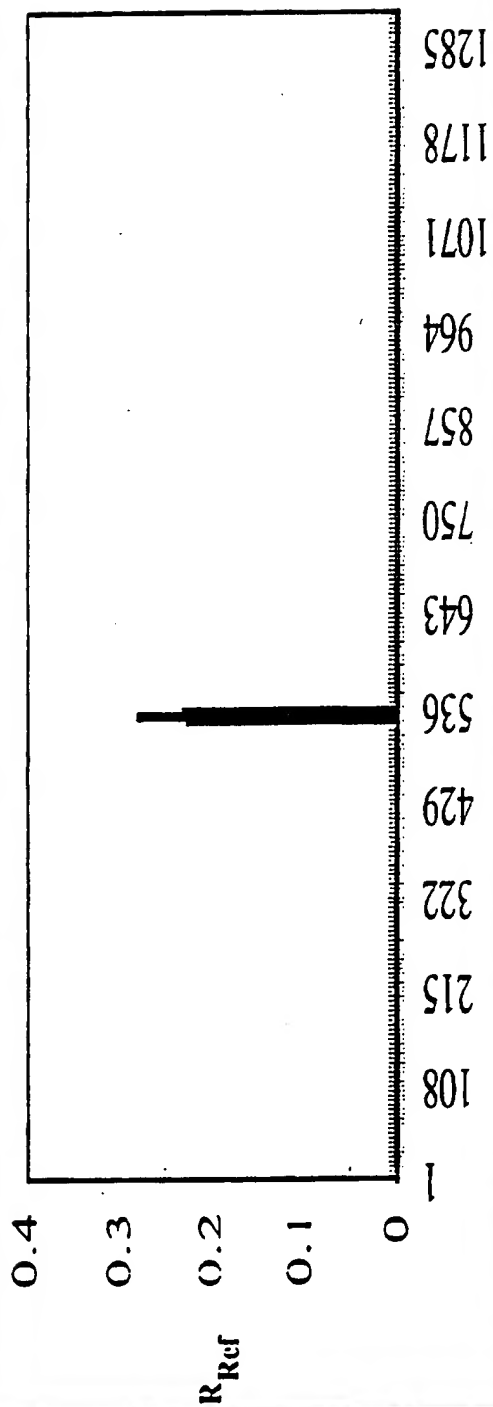


1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----

Figure 30

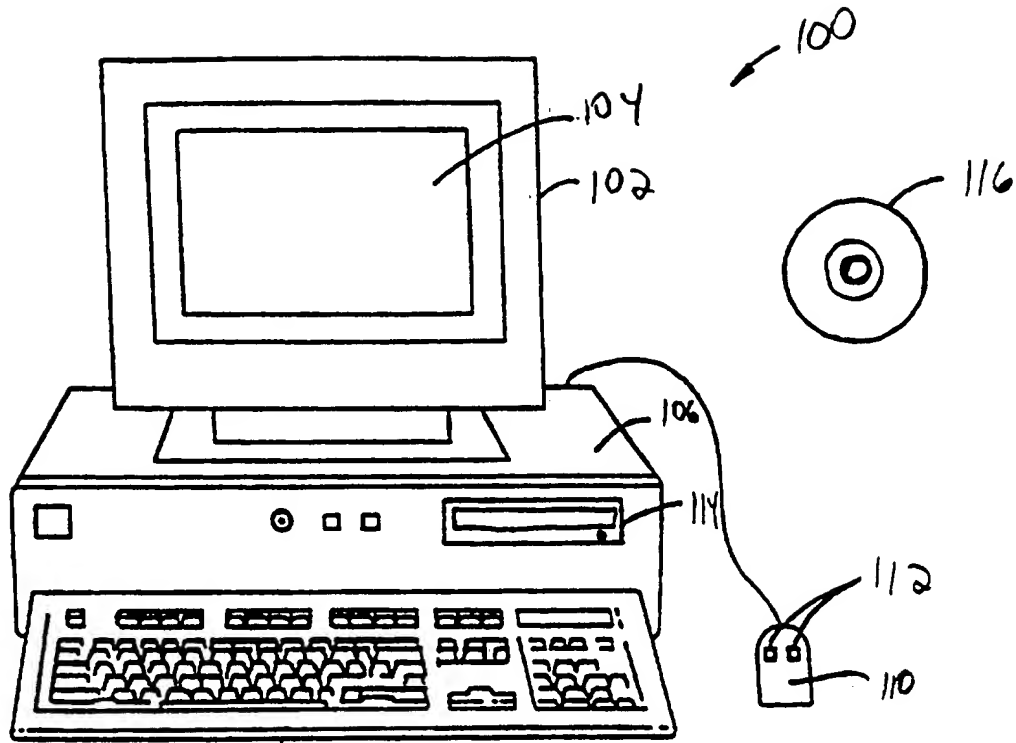
5000bp

T -> C bp substitution in Exon 5



2	3	4	5	6	7	8	9	10	11
---	---	---	---	---	---	---	---	----	----

Figure 31



108 FIG. 32

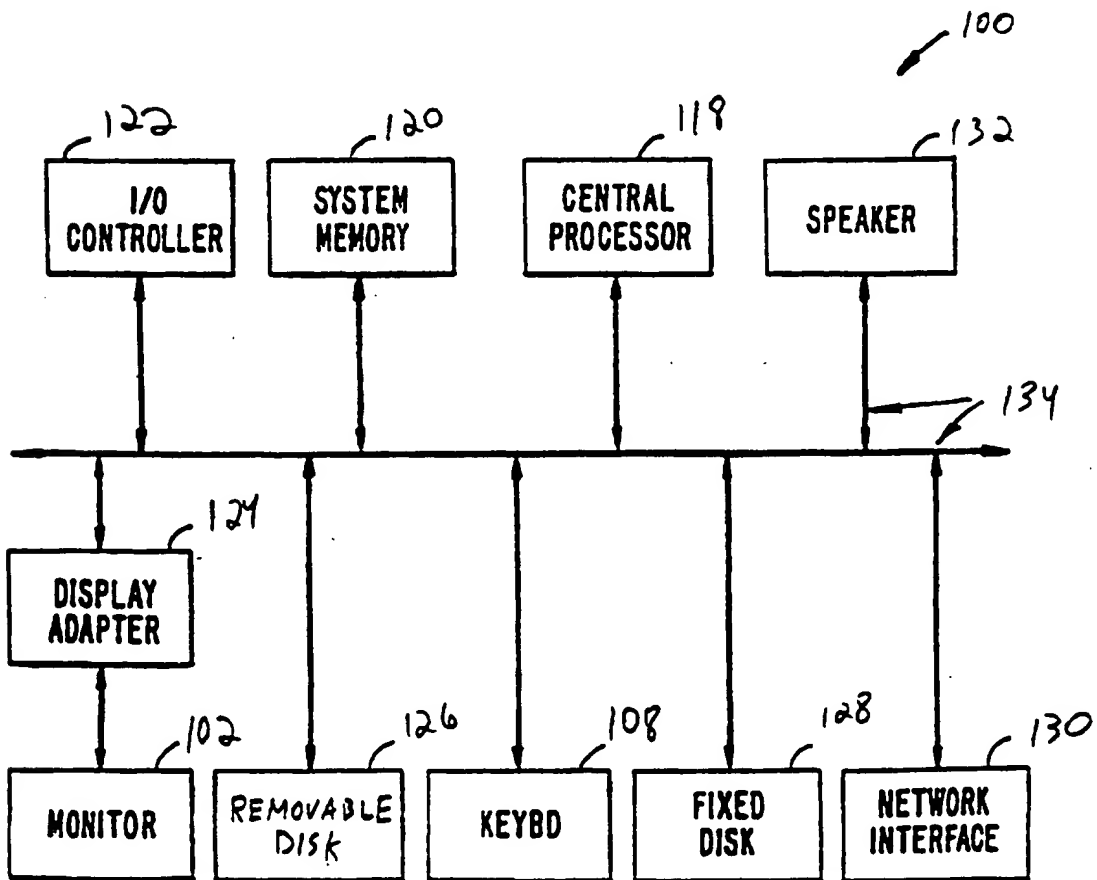


FIG. 33

33 / 36

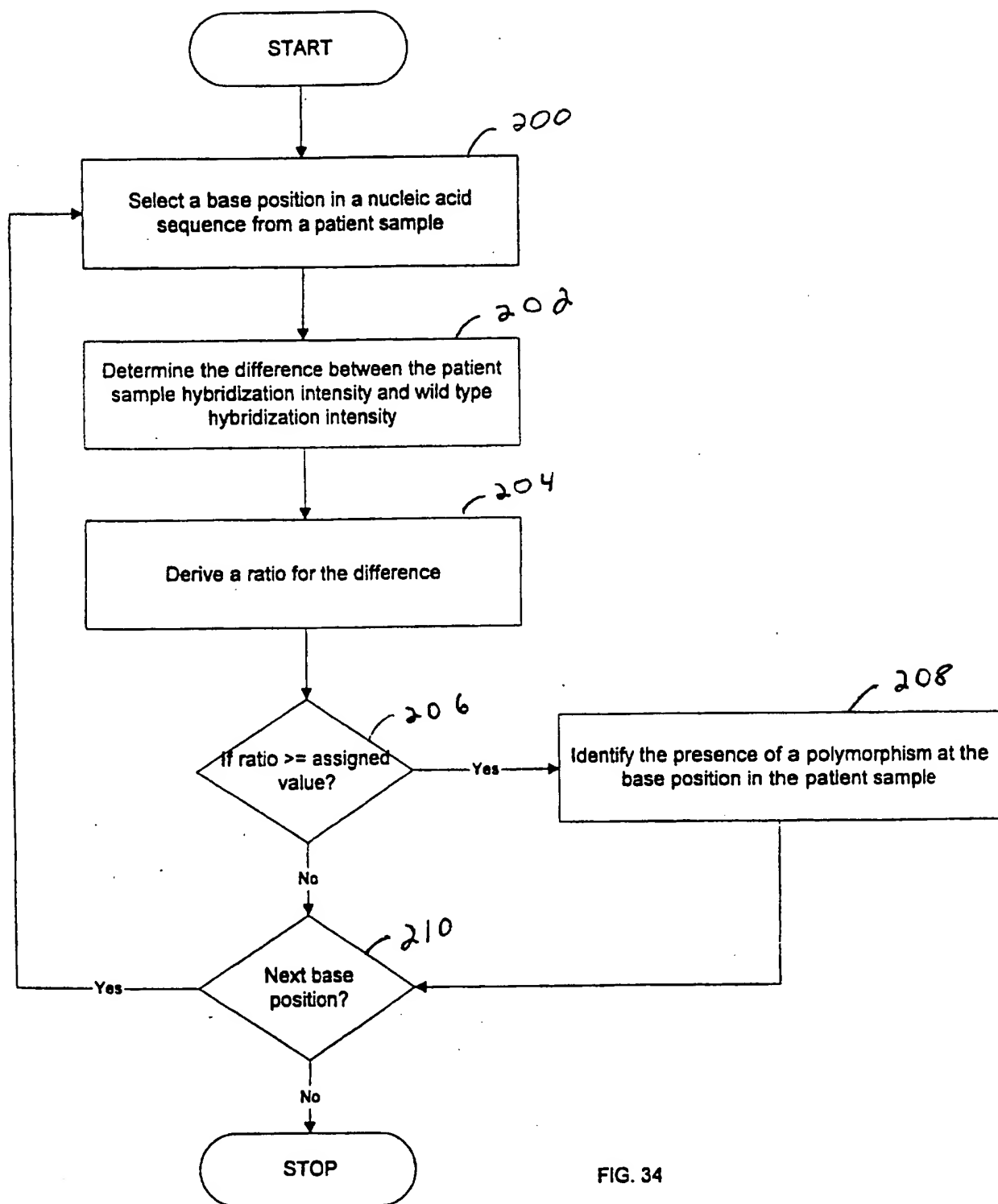


FIG. 34

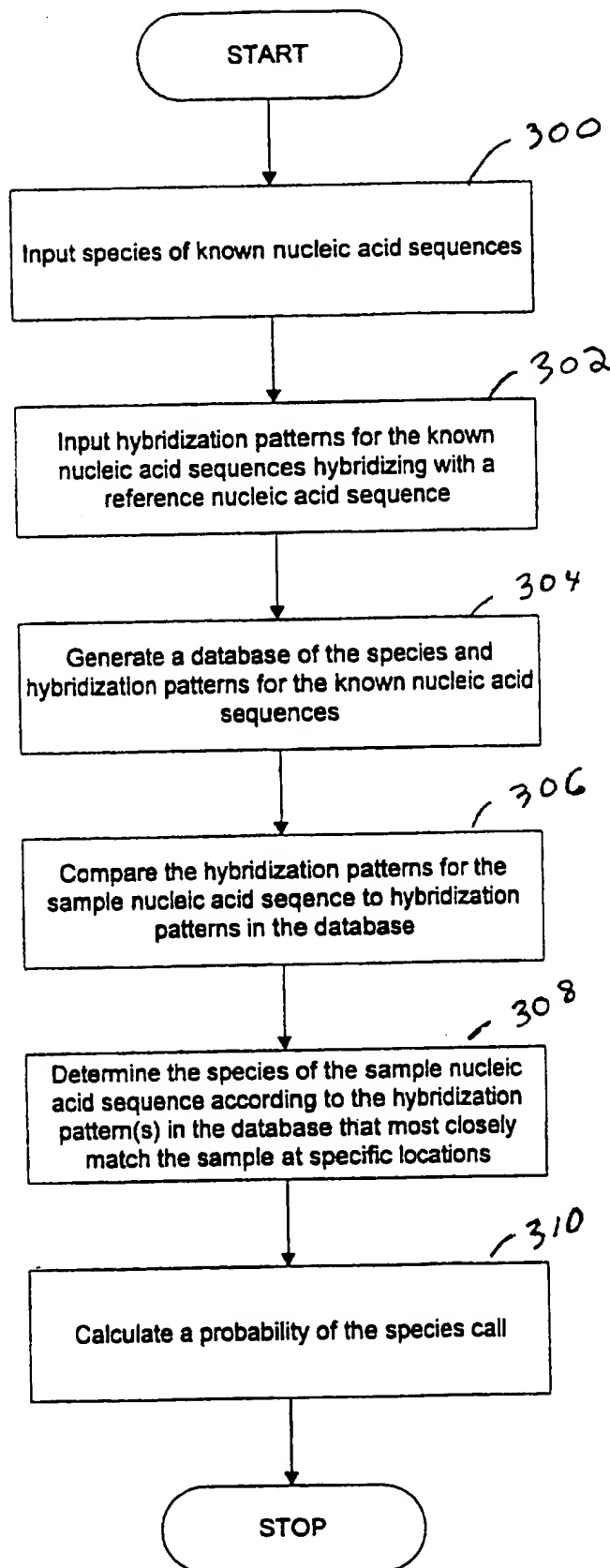


FIG. 35

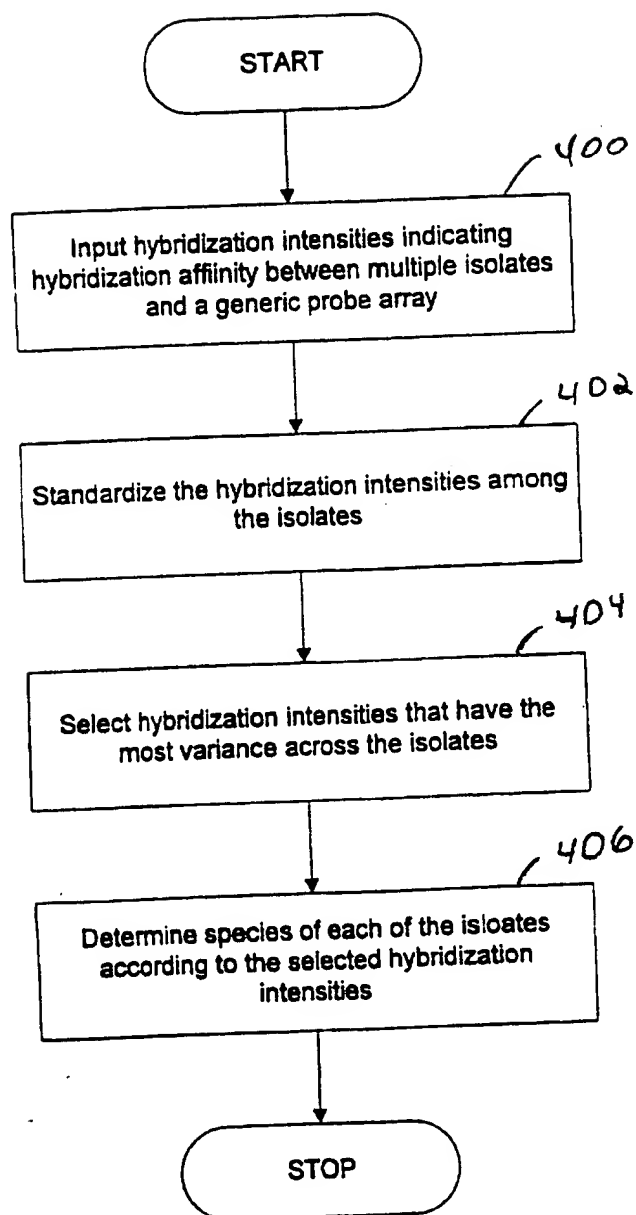


FIG. 36

3 6 / 3 6

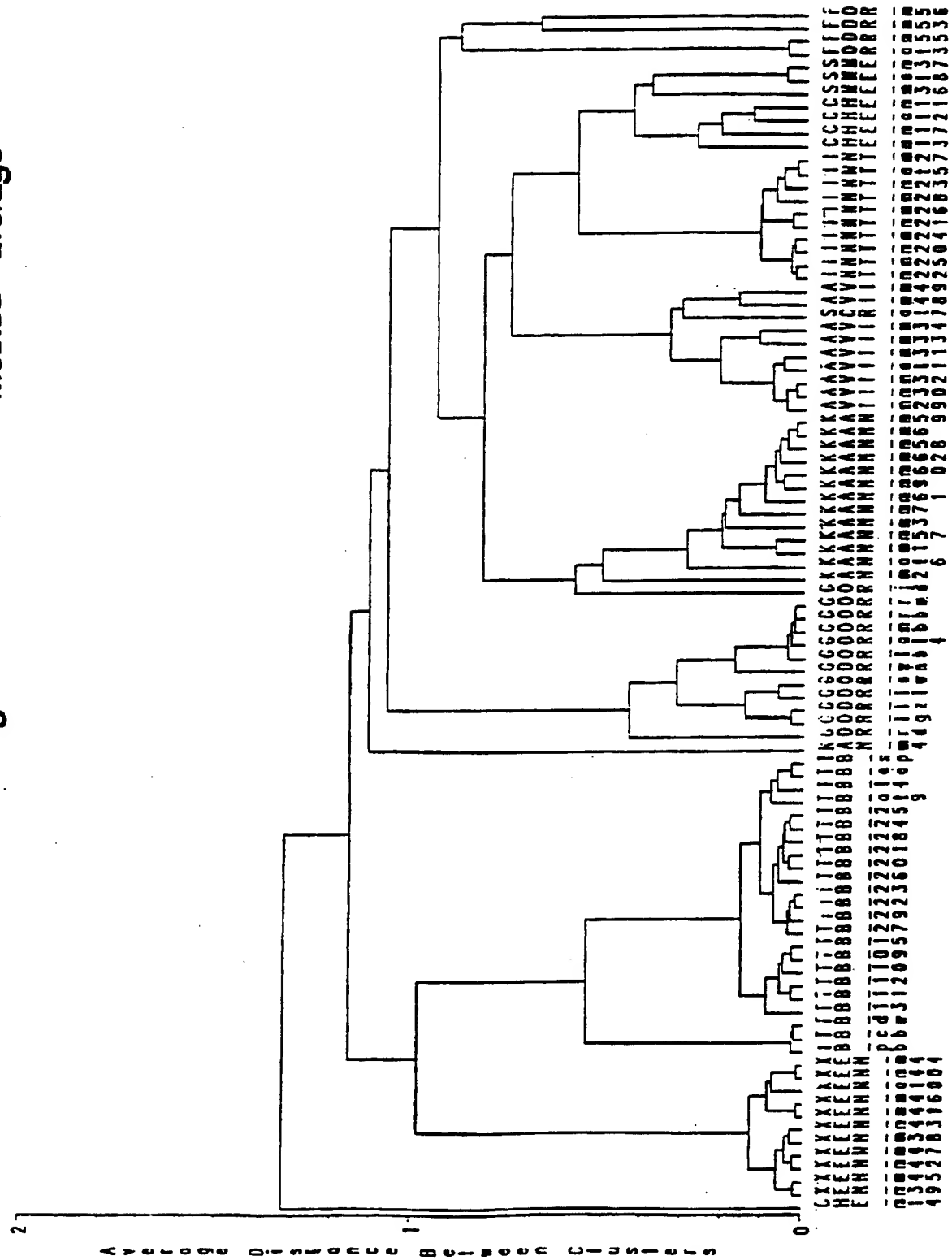


FIG. 37 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/02102

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68, 1/70; C07H 21/04; C12P 19/34

US CL : 435/5, 6, 91.2; 422/104; 536/24.3, 24.31, 24.32

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/5, 6, 91.2; 422/104; 536/24.3, 24.31, 24.32

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, CA, DERWENT

search terms: probe, oligonucleotide, array, mycobacteria, computer, algorithm, hybridization

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	LIPSHUTZ et al. Using oligonucleotide probe arrays to access genetic diversity. BioTechniques. September 1995, Vol. 19, No. 3, pages 442-447, especially pages 445-447.	1 ----- 2-89
X ---- Y	PEASE et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. USA. May 1994, Vol. 91, pages 5022-5026, especially page 5026.	1 ----- 2-89
Y	HUNT et al. Detection of a genetic locus encoding resistance to rifampin in mycobacterial cultures and in clinical specimens. Diagn. Microbiol. Infect. Dis. 1994, Vol. 18, pages 219-227, especially page 224.	2-5, 30-33, 53-56, 59-61, 74, 78, 80

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

22 APRIL 1997

Date of mailing of the international search report

13 MAY 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

KENNETH R. HORLICK

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/02102

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	WO 95/11995 A1 (AFFYMAX TECHNOLOGIES N.V.) 04 May 1995, see especially pages 19 and 20.	1-66 ----- 67-89

Detection of point mutations with a modified ligase chain reaction (Gap-LCR)

Klara Abravaya*, John J. Carrino, Sharon Muldoon and Helen H. Lee

Abbott Laboratories, Probe Diagnostics Business Unit, Abbott Park, IL 60064, USA

Received October 4, 1994; Revised and Accepted January 3, 1995

ABSTRACT

DNA amplification systems are powerful technologies with the potential to impact a wide range of diagnostic applications. In this study we explored the feasibility and limitations of a modified ligase chain reaction (Gap-LCR) in detection and discrimination of DNAs that differ by a single base. LCR is a DNA amplification technology based on the ligation of two pairs of synthetic oligonucleotides which hybridize at adjacent positions to complementary strands of a target DNA. Multiple rounds of denaturation, annealing and ligation with a thermostable ligase result in the exponential amplification of the target DNA. A modification of LCR, Gap-LCR was developed to reduce the background generated by target-independent, blunt-end ligation. In Gap-LCR, DNA polymerase fills in a gap between annealed probes which are subsequently joined by DNA ligase. We have designed synthetic DNA targets with single base pair differences and analyzed them in a system where three common probes plus an allele-specific probe were used. A single base mismatch either at the ultimate 3' end or penultimate 3' end of the allele specific probe was sufficient for discrimination, though better discrimination was obtained with a mismatch at the penultimate 3' position. Comparison of Gap-LCR to allele-specific PCR (ASPCR) suggested that Gap-LCR has the advantage of having the additive effect of polymerase and ligase on specificity. As a model system, Gap-LCR was tested on a mutation in the reverse transcriptase gene of HIV, specifically, one of the mutations that confers AZT resistance. Mutant DNA could be detected and discriminated in the presence of up to 10 000-fold excess of wild-type DNA.

INTRODUCTION

The ability to detect single base changes is of great importance in molecular genetics. Specific identification of point mutations in the human genome plays a major role in diagnosis of hereditary diseases and in identification of mutations within oncogenes, tumor suppressor genes and of mutations associated with drug resistance.

Single base variations have been analyzed by a variety of techniques, such as restriction fragment length polymorphism (1), denaturing gradient gel electrophoresis (2) and chemical cleavage of mismatched heteroduplexes (3). Other techniques include RNase cleavage of mismatched bases (4) and single strand conformation polymorphism (5). All of these techniques have the advantage of being able to screen for unknown mutations. Yet, they are very labor intensive, multistep, non-automated processes and most importantly lack sensitivity (6). Recently, highly sensitive amplification-based techniques have been developed, among which are hybridization of allele-specific oligonucleotides to polymerase chain reaction (PCR)-amplified products (7,8) and competitive oligonucleotide priming, where differential amplification depends on differential hybridization (9). The amplification refractory mutation system (10), also referred to as allele-specific PCR (ASPCR) (11), which relies on positioning the mutation at the 3' end of a PCR primer, and the ligase chain reaction (LCR), where a mismatch is positioned at the ligation joint (12-14), are two other amplification technologies used for analysis of single base mutations.

In the LCR, two pairs of synthetic oligonucleotides which hybridize at adjacent positions to complementary strands of a target DNA are joined by a thermostable ligase. Multiple rounds of denaturation, annealing and ligation result in the exponential amplification of the target DNA (Fig. 1A) (13-18). Targets that differ by a single base pair are discriminated, since a mismatch at the ligation joint severely reduces the efficiency of ligation (12-14,19). Generation of target-independent ligation products due to blunt-end ligation poses limitations on the sensitivity of LCR (20). Typically, the sensitivity of LCR or any diagnostic assay is not a critical factor for detection of mutations in human genetic diseases, where 50 or 100% of DNA contains the mutation. In contrast, for detection of somatic mutations within oncogenes, tumor suppressor genes or drug resistance mutations, where a small number of mutated molecules need to be detected in the presence of excess wild-type DNA, sensitivity becomes a critical factor.

Several approaches have been taken to increase the sensitivity of LCR. One approach has been to use another amplification technology, such as PCR, followed by limited amplification with LCR (21,22). Other alternatives are PCR followed with the ligation detection reaction (LDR), where only two adjacent probes are used, resulting in linear amplification (13,14,20,21), or PCR followed with the oligonucleotide ligation assay (OLA),

* To whom correspondence should be addressed

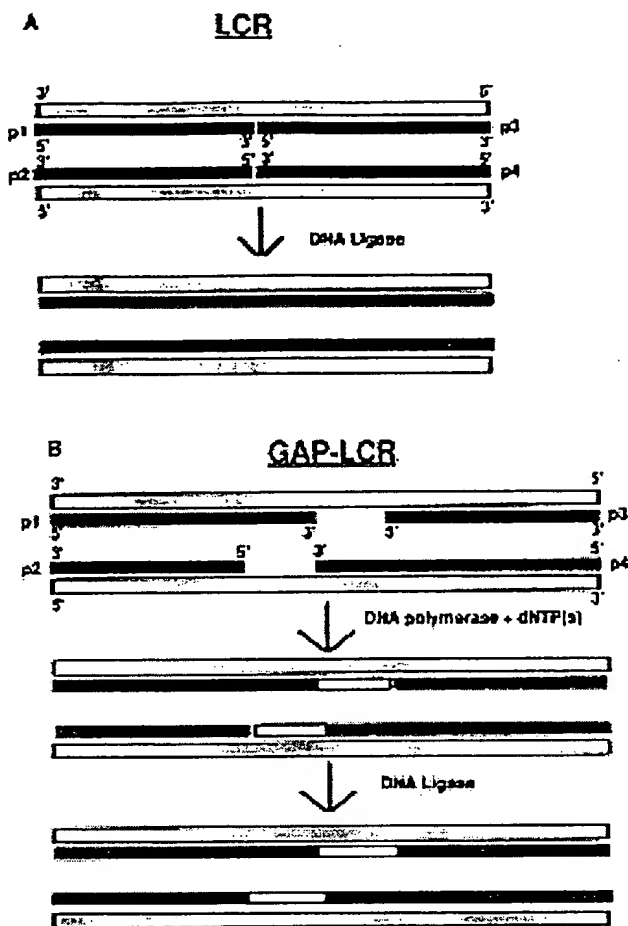


Figure 1. Diagrammatic representation of LCR and Gap-LCR. The complementary strands of target DNA are represented as shaded bars, LCR probes as solid bars and regions extended by DNA polymerase as white bars. (A) In LCR, four probes covering the entire target sequence anneal to complementary strands, probe 1 is ligated to probe 3 and probe 2 to probe 4 by a thermostable DNA ligase. The ligated probes function as targets in subsequent cycles and exponential amplification is achieved. (B) In Gap-LCR, probes 1 and 4 have 3' overhangs with respect to their complements. Probes 1 and 4 are extended by DNA polymerase with the appropriate nucleotide(s) to fill the gap and ligated to probes 3 and 2 respectively with the ligase.

where ligation of two adjacent probes is used as a single detection step (12,19,23). However these combined approaches necessitate the opening of tubes after PCR, generating a source of contamination and also introducing complexity to automation.

A modification of LCR, Gap-LCR has been introduced to circumvent these difficulties and improve the sensitivity of LCR (24–27). In Gap-LCR, complementary probe pairs containing 3' extensions are used. After hybridization to target DNA, a gap of one to several bases exists between adjacent probes. A thermostable DNA polymerase, devoid of 3'→5' exonuclease activity, and the appropriate nucleotide(s) are used to fill the gap and the resultant probes are joined by DNA ligase (Fig. 1B). The use of probe duplexes with non-complementary 3' extensions prevents the generation of target-independent ligation products. Gap-LCR has been successfully applied to detect < 10 target molecules in a reaction (unpublished results). Amplification products are detected by a sandwich immunoassay performed with an

automated analyzer (17,26). The sensitivity, specificity and automation of the technology make Gap-LCR a good candidate for diagnostic tests.

In this study we explored the properties of Gap-LCR in the detection and discrimination of target DNA sequences that differ by a single base pair.

MATERIALS AND METHODS

Oligonucleotides and plasmids

Target DNAs (50 nt) were synthesized and gel purified by Genosys (The Woodlands, TX). The sequence of the wild-type target was derived from the sequence of the *Chlamydia trachomatis* cryptic plasmid, map position 2230–2280 (28). Mutant A and Mutant B targets were identical to the wild-type target, except single base changes were introduced at the indicated positions to both strands during synthesis (Fig. 2A). Targets for HIV experiments (Fig. 5) were gifts from Dr Steve Wolinsky (Northwestern University) and Dr John Mellors (University of Pittsburgh). They were provided as purified DNA from plasmids containing a 1.7 kb fragment of the HIV genome cloned into *Eco*R1 and *Hind*III sites of the vector pKK233 (Pharmacia). The mutant sequence has a mutation at amino acid 215 which changes the codon from ACC to TAC (from threonine to tyrosine) (Fig. 5A). The sequence of the region used as the target for LCR is 5'-AACATCTGTTGAGGTGGGGATTTACCACACACCAGACAAAAACATCAGA. The LCR probe sets were synthesized on an Applied Biosystems synthesizer 394 by the phosphoramidite method. The 5' end of probe 1 and 3' end of probe 2 were covalently linked to carbazole, while the 3' end of probe 3 and 5' end of probe 4 were linked to adamantane (Fig. 2A). In the experiments where both PCR and LCR were performed (Fig. 4), only probes 1 and 4 were haptenated. Probes were purified on a 12% denaturing polyacrylamide gel (29). Quantitation was by absorbance at 260 nm.

LCR and PCR amplification

LCR and PCR reactions contained 500 ng human placental DNA with either no target DNA (negative control) or with 100 molecules of target DNA unless stated otherwise. LCR reactions were run in a buffer containing 50 mM EPPS, pH 7.8, 30 mM MgCl₂, 20 mM K⁺, 10 μM NAD, 1–10 μM gap filling nucleotides, 30 nM each oligonucleotide probe, 1 U *Thermus flavus* DNA polymerase, lacking 3'→5' exonuclease activity (MBR, Milwaukee, WI), and 5000 U *T.thermophilus* DNA ligase (Abbott Laboratories; 1 U is the amount of DNA ligase producing 1 nM ligated product in 10 min at 55°C at pH 7.8). Reaction volume was 50 μl and each reaction was overlaid with 50 μl mineral oil prior to cycling in a Perkin Elmer 480 thermocycler. Cycling conditions consisted of a 30 s incubation at 85°C and a 30 s incubation at 60°C. Cycle numbers are indicated in figure legends. PCR reactions were run under the same conditions as LCR, except all four dNTPs were used and probes 2 and 3 and ligase were omitted. For the amplification of HIV sequences, concentrations of LCR reagents were as described above except 0.5 U DNA polymerase was used. The reaction cycling conditions consisted of a 3 min denaturation at 94°C followed by 38 cycles of 1 s at 94°C, 1 s at 58°C and 30 s at 64°C.

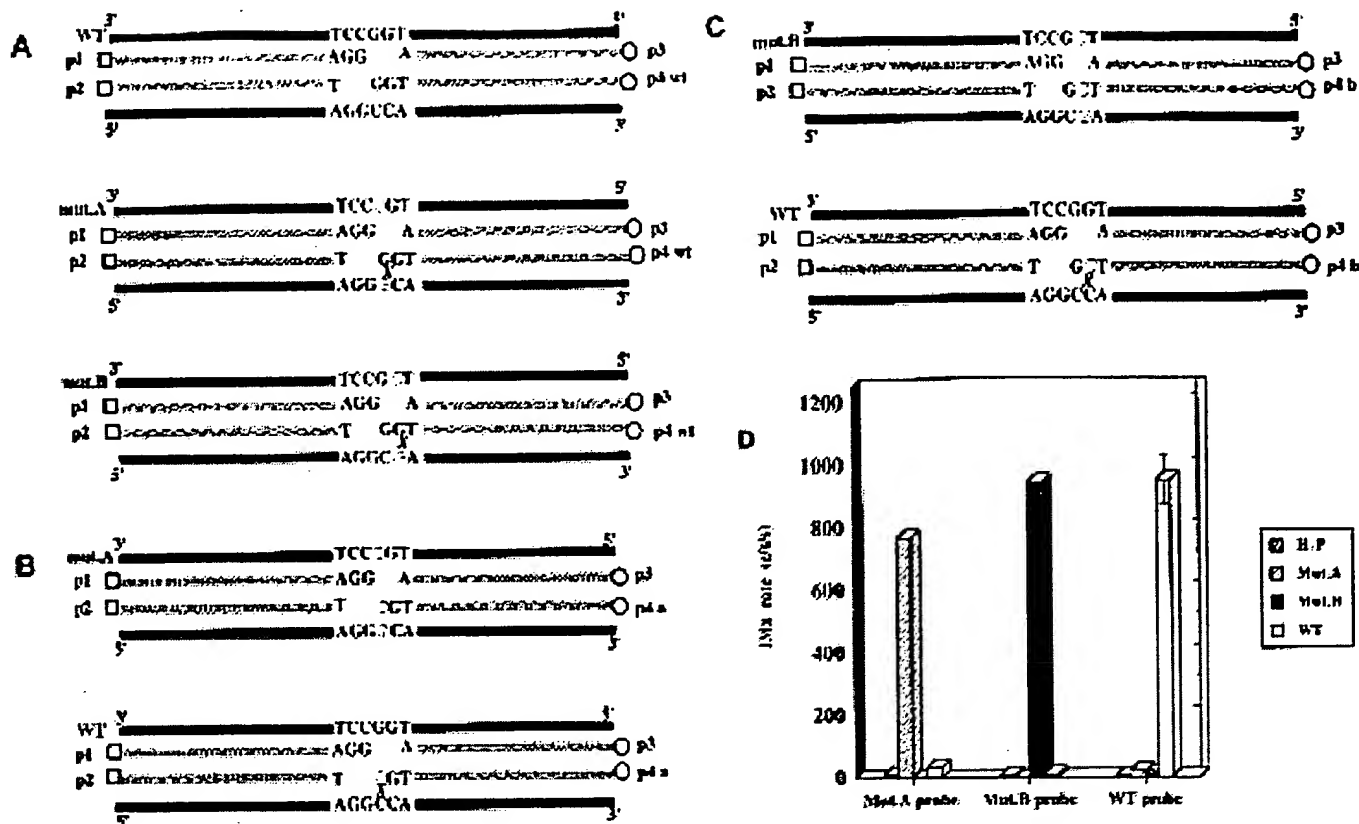


Figure 2. Design and specificity of Gap-LCR probe sets for the amplification of DNAs that differ by a single base pair. Double-stranded synthetic target DNAs, designated wild-type (WT), mutant A or mutant B are shown. Only nucleotides of interest are shown; the remainder of the sequences represented by black bars are identical in all targets. The changed nucleotides in the mutant A and mutant B targets with respect to the wild-type target are highlighted. Gap-LCR probes are represented as gray bars. Carbazole is represented as squares and adamantane as circles. (A) Gap-LCR probes for the specific amplification of the wild-type target. Probe 4 is complementary to the wild-type target and has a single mismatch (X) with mutant A and mutant B targets. (B) Gap-LCR probes for the specific amplification of the mutant A target. Probes 1, 2 and 3 are the same as in (A), probe 4a is different from probe 4wt and the change is highlighted. P4a is complementary to the mutant A target and has a mismatch (X) with the wild-type target. (C) Gap-LCR probes for specific amplification of the mutant B target. The change in probe 4b is highlighted. P4b is complementary to the mutant B target and has a mismatch (X) with the wild-type target. (D) Specificity of the Gap-LCR probes. The wild-type-, mutant A- and mutant B-specific probe sets shown in (A-C) respectively were tested with either human placental DNA (H.P.), wild type (WT) target, mutant A or mutant B targets. Reaction conditions are described in Materials and Methods. Samples were cycled for 25 cycles and products were detected using the Abbott IMx[®] automated immunoassay as counts/second/second (c/s/s) as described (17).

Detection of amplified products

Amplification products were detected via a sandwich immunoassay performed using the Abbott IMx[®] automated analyzer. Amplification products were captured using anti-carbazole coated microparticles. After a washing step, the captured products were detected using an anti-adamantane-alkaline phosphatase conjugate which, in the presence of methylumbelliferone phosphate, generates a fluorescent product at a rate proportional to the amount of captured product. The average IMx[®] rate from duplicate samples was taken and standard deviations are shown.

For the detection of LCR products on polyacrylamide gel (Fig. 3B), unphosphorylated probe 1 was phosphorylated at the 5' end using the Gibco BRL 5' DNA terminus labeling system and 50 μ Ci [γ -³²P]ATP (Amersham). LCR reactions were set up as described above, except that equal amounts of radiolabeled and cold probe 1 were used (15 nM of each per reaction), and samples were cycled for 43 cycles. For restriction analysis, 15 μ l of the amplified product was incubated with 1.5 μ l *Hae*III (10 U/ μ l) and 1.8 μ l 10 \times buffer (Promega) for 1 h at 37°C; the controls (–) lanes were also incubated with 10 \times buffer at 37°C in the absence of

*Hae*III. Products were separated by electrophoresis on 12% denaturing polyacrylamide gels (29).

RESULTS

Design of targets and Gap-LCR probe sets

Synthetic double-stranded DNA targets, designated wild-type or mutant, that differed by a single base pair were designed as shown in Figure 2A. Gap-LCR probe sets specific for the amplification of each target DNA were synthesized. The probes were staggered, i.e. probes 1 and 4 had 3' overhangs when hybridized to their complements. When annealed to the target DNA, probes 1 and 4 were extended by DNA polymerase, in the presence of appropriate nucleotide(s). The extended probes were then ligated to probes 3 and 2 respectively. The ligated products could function as targets in subsequent cycles, thus allowing exponential amplification. Probes 1, 2 and 3 were common and probe 4 was specific for each target; probes 4wt, 4a and 4b were designed to specifically amplify wild-type, mutant A and mutant B targets respectively and had a single mismatch with the non-analogous

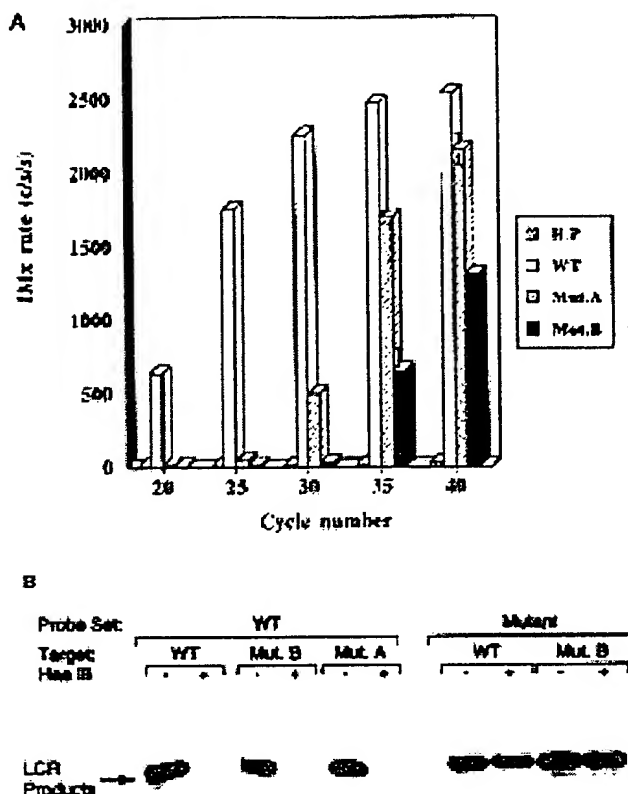


Figure 3. (A) The effect of cycle number on specificity of Gap-LCR. the wild-type-specific probe set was used with human placental (H.P.), wild-type (WT), mutant A and mutant B DNAs as shown in Figure 2A. Reaction conditions are as described in Materials and Methods and cycle numbers are indicated. (B) Analysis of 'overamplified' LCR products. the wild-type-specific probe set was tested with wild-type, mutant B and mutant A targets as shown in Figure 2A. the mutant B-specific probe set was tested with wild-type and mutant B target as shown in Figure 2C. Probe 1 was radiolabeled with ^{32}P and reactions were cycled for 43 cycles. Amplified products were divided in two; one set was restricted with *HaeIII* (+ lanes), the other set was not (- lanes). Products were electrophoresed on a 12% denaturing gel and detected by autoradiography.

targets (Fig. 2A-C). The mutation was positioned so that on one strand it was complementary to one of the bases to be filled during the extension of probe 1 and on the other strand it was mismatched with respect to probe 4. To assess the effect of the mismatch position on specificity, the position of the mutation was varied to generate a C:C mismatch either at the ultimate 3' end or penultimate 3' end of probes 4a and 4b with respect to the wild-type target (Fig. 2B and C). Similarly the wild-type-specific probe, 4wt, had a G:G mismatch either at the ultimate 3' end or penultimate 3' end with mutant A and mutant B targets respectively (Fig. 2A).

Specificity of Gap-LCR probe sets

The specificity of the probe sets is shown in Figure 2D. Mutant A, mutant B, wild-type or human placental DNA (negative control) were amplified with the different probe sets. With human placental DNA, amplified product was not observed, indicating that target-independent non-specific amplification was not significant. Mutant probe sets amplified only their respective

mutant targets, whereas the wild-type probe set amplified only the wild-type target and not the mutant targets. These results demonstrate that a single base mismatch positioned either at the ultimate 3' or penultimate 3' end of probe 4 is sufficient to provide discriminative amplification by Gap-LCR under the conditions used in this study.

As has been shown for ASPCR and LCR (30), discrimination by Gap-LCR can be adversely affected by increasing the number of amplification cycles. To determine the maximum number of cycles where the amplification remains specific, wild-type, mutant A, mutant B and human placental DNAs were amplified with the wild-type-specific probe set in the presence of dCTP for 20, 25, 30, 35 or 40 cycles (Fig. 3A). Amplified product was detected after 20 cycles with wild-type target, after 30 cycles with mutant A target and after 35 cycles with mutant B target. No product was detected with human placental DNA even after 40 cycles. This result suggests that there is a window of about 10 cycles where the amplification is most specific. Similar results were observed when mutant-specific probes were used with each target (data not shown).

When the wild-type probe set was used with mutant targets (Fig. 2A), identical specificity was seen when dGTP was omitted or added to the reaction, suggesting that omission of the nucleotide to fill the base complementary to the mutation does not significantly contribute to specificity (data not shown). This result was expected, since extension of probe 1 with dCTP and dGTP and ligation to probe 3 would not generate a perfect substrate for probes 2 and 4; probe 4 would still be mismatched with the ligated substrate and be refractory to amplification (Fig. 2A). In contrast, extension from the mismatched probe 4 and ligation to probe 2 would generate a ligated product that would be a perfect substrate for probes 1 and 3, in which case dGTP would not be needed and wild-type product would be generated (Fig. 2A). This prediction was confirmed experimentally by analyzing the products that were generated after over-amplification. wild-type and mutant targets were amplified with the wild-type probe set for 43 cycles (where products from mismatched targets are generated) in the presence of both dGTP and dCTP and the nature of the amplified products was analyzed (Fig. 3B). Amplified products were digested with the restriction enzyme *HaeIII*, which cleaves at the GGCC site which would be present only on wild-type products (Fig. 2A). For this experiment, the 5' end of probe 1 was radiolabeled and the products were detected on a denaturing polyacrylamide gel. The results demonstrate that the products generated from both matched and mismatched targets were cleaved by *HaeIII*, thus wild-type product was generated in all cases (Fig. 3B). In contrast, products amplified with the mutant B-specific probe set were not cleaved by *HaeIII*. Products generated with the mutant A-specific probe set were not cleaved by *HaeIII* either (data not shown). These results confirm the prediction that dGTP is not utilized in the generation of products when the wild-type probe set is used with mutant targets. Therefore omission of dGTP does not significantly contribute to the specificity.

We explored the specificity of Gap-LCR with increasing number of target molecules to determine the maximum number of mismatched target molecules where the amplification remains specific. the wild-type probe set was tested with increasing concentrations of matched (wild-type) or mismatched (mutant) targets (Fig. 4A). The results indicate that while 10 molecules of the matched target were detected, using optimal cycle numbers,

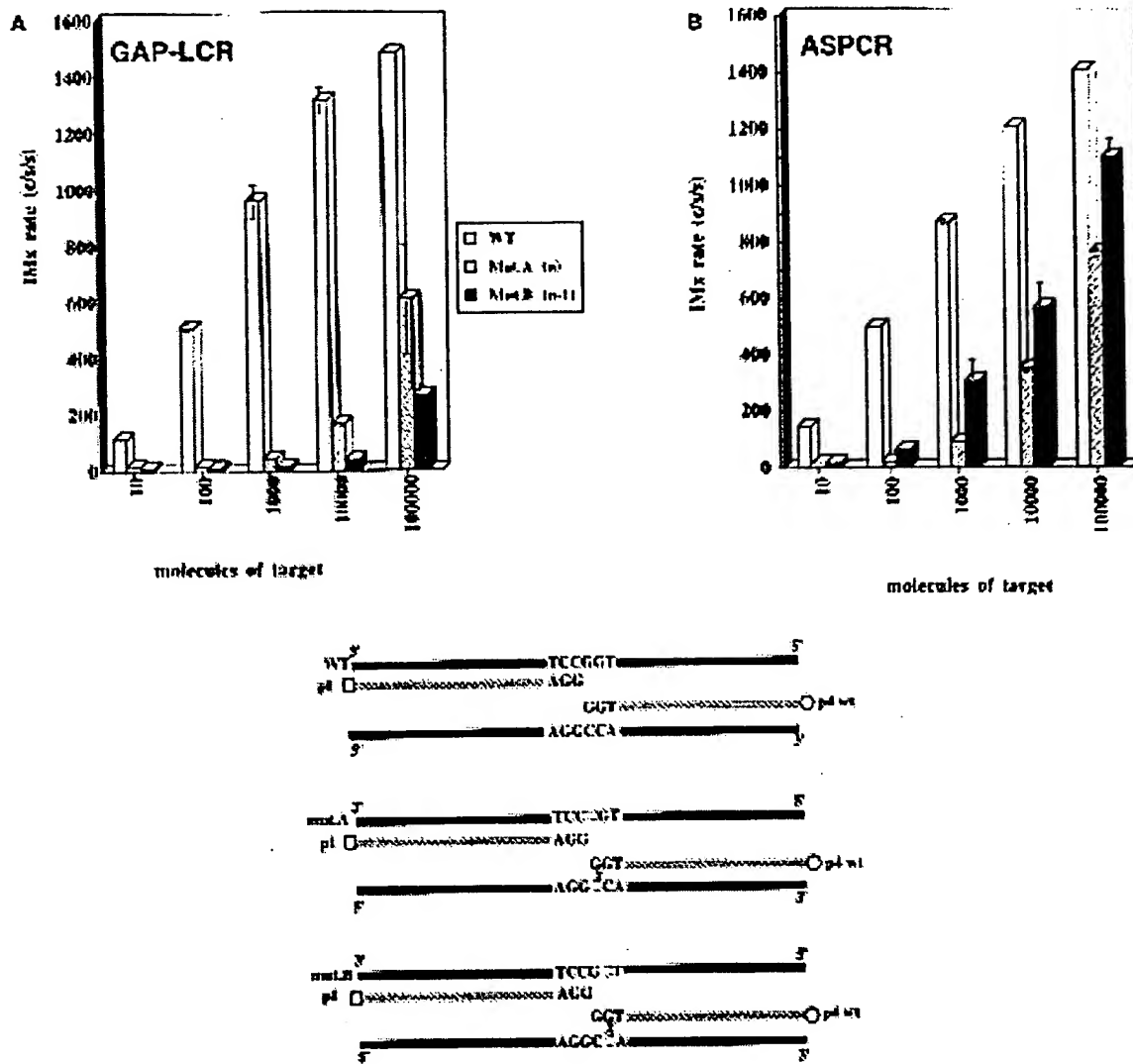


Figure 4. Comparison of Gap-LCR to ASPCR. (A) The wild-type-specific Gap-LCR probe set was tested with wild-type, mutant A and mutant B targets as shown in Figure 2A, except only probes 1 and 4 were haptenated. Target concentrations are as shown. (B) PCR reactions were as described in Materials and Methods; reactions were cycled for 23 cycles. The primers used for PCR are shown in the lower panel. They are identical to probes 1 and 4 used in Gap-LCR (Fig. 2A). The primers are specific for amplification of the wild-type target. Primer 4 has a mismatch at the ultimate 3' end with the mutant A target and at its penultimate 3' end with the mutant B target (Xs).

detection of the mismatched targets occurred only with 10^4 – 10^5 molecules. The loss of discrimination with mutant A target (10^4 molecules) preceded the loss of discrimination with mutant B target (10^5 molecules). Similar results were seen when the cycle number was increased beyond the optimum (Fig. 3A); product was detected after 30 cycles with mutant A target and 35 cycles with mutant B target. Both mutant targets amplified at the same rate with their respective matched probes and the differential rate of amplification of mutant targets was only observed with mismatched wild-type probes. These results suggest that a mismatch positioned at the penultimate 3' end is discriminated better than a mismatch at the ultimate 3' end.

In Gap-LCR, discrimination between targets that differ by a single base may rely on three steps: (i) hybridization of mismatched probes; (ii) fidelity of the polymerase to extend from mismatches; (iii) specificity of the ligase to join probes extended from mismatches. ASPCR also requires the first two steps, yet

Gap-LCR may have the additional level of specificity required by the necessity for proper ligation. To address this question, specificity of Gap-LCR and ASPCR were compared under the same reaction conditions (Fig. 4). ASPCR experiments were performed with the same targets using only two of the haptenated probes (probes 1 and 4), all four nucleotides and same reaction conditions utilized for Gap-LCR. For this comparative study, only probes 1 and 4 were linked to haptens for Gap-LCR. Detection of products relied on complementarity of the strands linked to the two haptens. Results indicate that with ASPCR, better discrimination was observed when the mismatch was at the ultimate 3' end than at the penultimate 3' end. This is in contrast to the observation made using Gap-LCR (Figs 3A and 4A). Comparing the two amplification procedures, mismatched targets were amplified at a faster rate in ASPCR than Gap-LCR, while the amplification rate of the matched target was equivalent in both reactions. This difference was enhanced when the mismatch was

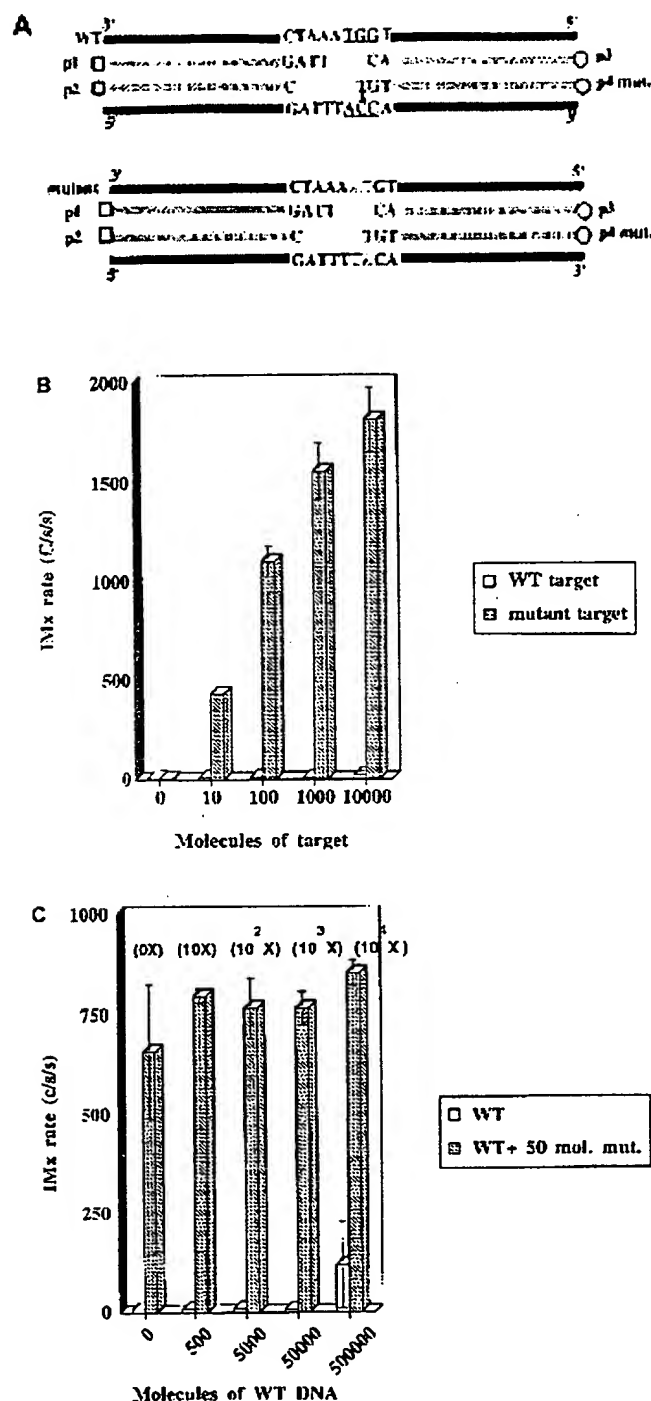


Figure 5. Gap-LCR for the detection of an AZT resistance mutation. (A) Design of the Gap-LCR probe set for specific amplification of the mutant DNA. The wild-type and mutant target DNAs comprising 50 bases of the HIV reverse transcriptase gene are shown as solid bars. Codon 215 is underlined. The mutations in codon 215 are highlighted. Probes are represented as gray bars and haptenated as described in Figure 2. Probe 4mut. is specific for amplification of the mutant DNA and has a 3' terminal mismatch (X) with the wild-type target. (B) The mutant-specific probe set was tested with increasing concentrations of wild-type or mutant targets as shown in (A). Reaction conditions are as described in Materials and Methods. (C) The mutant-specific probe set was tested with 50 molecules of mutant target mixed with increasing concentrations of the wild-type target (shaded bars) or with increasing concentrations of the wild-type target alone (open bars). The ratio of the wild-type target to the mutant target is indicated above the shaded bars. Reaction conditions are as described in Materials and Methods.

at the penultimate 3' end. This observation suggests that the specificity of Gap-LCR does not solely rely on hybridization and the extension of the mismatched probe by the polymerase as in ASPCR and that the ligation step adds to the specificity of Gap-LCR.

Gap-LCR for detection of a HIV AZT resistance mutation

To show the feasibility and specificity of Gap-LCR for detection and discrimination of a natural mutation, a mutation at codon 215 of the HIV reverse transcriptase gene was tested as a model system. Mutation at codon 215 from ACC (threonine) to TAC (tyrosine) has been associated with resistance to AZT (3'-azido-3'-deoxythymidine) (31). Probe sets specific for the amplification of the mutant viral DNA were designed and tested on cloned HIV DNA carrying wild-type or mutated sequences at codon 215 (Fig. 5A). Even though the wild-type and mutant targets differ by two bases, one of them (the first base of the codon) is positioned in the overlapping gap and does not create a mismatch with the probes. Because the necessary nucleotides are provided in the reaction (ATP and TTP), the change in that position does not contribute to discrimination. Thus, the discrimination relies on a single base change; the mutant-specific probe set is designed to have a single mismatch at the ultimate 3' end of probe 4 with the wild-type DNA. The specificity of the mutant-specific probe set was tested with increasing concentrations of wild-type and mutant targets. While 10 molecules of mutant target were detected, no product was observed with up to 10 000 molecules of wild-type target (Fig. 5B). No product was detected with human placental DNA (used as a negative control). To determine whether the mutant target could be detected in the presence of excess wild-type target, 50 molecules of mutant target were mixed with increasing concentrations of wild-type target (5×10^2 to 5×10^5). Increasing concentrations of wild-type target in the absence of mutant target was tested as the control. Results indicate that specific amplification of the mutant target occurs in the presence of up to 10^4 -fold excess wild-type target (Fig. 5C). Slight cross-reactivity with the wild-type target was observed when 5×10^5 molecules of wild-type target were used.

DISCUSSION

A reliable DNA diagnostic method requires accurate discrimination, low background and automation. In this study we have shown that Gap-LCR meets these requirements; DNA targets that differ by a single base pair were discriminated, the background was low, sensitivity was high and the products of the reaction were detected by an automated immunoassay. In the experiments designed in this study, discrimination between related targets relied on a single base mismatch between one of the Gap-LCR probes and the target DNA. The extent of the discrimination depended on the cycle number, concentration of the mismatched target and the position of the mismatch.

As with any other amplification reaction, the reaction specificity was expected to deteriorate with increasing cycle number (30). When mismatched probes are extended and ligated during any cycle, the newly formed molecules are able to function as templates in subsequent cycles. The products generated from the matched target will reach a plateau after a certain cycle number, while the products generated from the mismatched target will continue to exponentially amplify until they reach the levels seen

with the matched target, at which point discrimination will be completely lost. A rough calculation of the accumulation of products from matched and mismatched targets has been reported previously for ASPCR by Ugozoli and Wallace (30). We show that for Gap-LCR accumulation of products from mismatched targets occurs about 10–15 cycles later than the detection of products from the matched target. Moreover, even after 20 additional cycles (40 cycles total), no signal was observed with negative control placental DNA.

Template concentration also plays a major role in specificity. Similarly to an increased number of cycles, the specificity of Gap-LCR was expected to deteriorate when large amounts of mismatched target were used. We showed that under the conditions used in this study, as few as 10 molecules of the matched target were detected, while equivalent detection of the synthetic mismatched target required 10 000–100 000 molecules, depending on the position of the mismatch. With the HIV-specific probe set, the specificity was even better. Product was detected with 10 molecules of the matched target, while only a small amount of product was detected with 500 000 molecules of the mismatched target. The better specificity observed with the HIV probe set may be attributed to differences in sequence, in reaction conditions and/or the nature of the targets used in these studies (50 bp linear double-stranded synthetic targets versus 8 kb circular plasmid DNA). Our results indicate that with Gap-LCR, good specificity can be obtained under conditions where exquisite sensitivity is maintained.

Our studies demonstrate that a single mismatch between one of the Gap-LCR probes and the target is sufficient for discrimination of single base substitutions. The specificity seems to rely solely on the efficiency of extension and ligation of the mismatched probe. Omission of the nucleotide complementary to the mutated base in the fill does not significantly add to the specificity. Once the mismatched probe is extended and ligated, it generates a target for the complementary probes, which can extend and ligate in the absence of the omitted nucleotide. After such an event, amplification is exponential in the following cycles. The omission of the nucleotide complementary to the mutation would effectively prevent amplification if the mutation was positioned in an overlapping gap, where probes need to be designed not to cover the mutated base in either strand. However, such a scheme has limited application. It can only be used if the mutation is an A or T change to a C or G or vice versa. Other changes would necessitate the same fill in the overlapping gap.

In several previous reports where single base mismatches were not refractory to amplification, further deliberate mismatches were introduced to achieve discrimination with ASPCR or blunt-end LCR (10,11,30,32–35). For ASPCR, Newton *et al.* (10) reported that the primers became increasingly refractory to amplification as the additional mismatch was moved progressively closer to the 3' end of the PCR primer. Under the conditions used in our study, a second mismatch was not necessary and in fact positioning a second mismatch next to the terminal mismatch would likely result in a failure to amplify either target, since we demonstrated that a single base mismatch one base from the 3' end was inhibitory to amplification. We have also observed that a mismatch two bases from the 3' end was refractory to amplification with Gap-LCR (data not shown). Reaction conditions may be optimized to accommodate additional mismatches. Whether such an approach would increase the specificity of Gap-LCR remains to be explored.

It was previously reported that the nature of the mismatch affects both the polymerase extension and ligation efficiencies (10,14,32,33). However, in ASPCR conflicting results were obtained for the same mismatches in different reports, presumably due to differences in primer length, surrounding sequences and reaction conditions (30). Although our studies were not designed to compare the effect of base pair composition on the specificity of Gap-LCR, we observed that G:G, C:C and C:T mismatches were all refractory to amplification with Gap-LCR. Nevertheless, to assess the effect of mismatch position on Gap-LCR specificity, we chose to limit our comparison to the same mismatch (G:G), positioned at two different locations at the ultimate or penultimate 3' end of probe 4wt. We found that better specificity was obtained when the mismatch was at the penultimate 3' end. Previous studies on the effect of mismatch positioning on the specificity of polymerases or ligases are limited in scope. The effect of mismatches for *Taq* polymerase and T4 ligase have been shown to be greatest at the ultimate 3' position (32,36). Our PCR results with *T.flavus* polymerase are in agreement with these observations. Yet in Gap-LCR we observed better specificity with the 3' penultimate mismatch. This difference may be due to the nature of the ligase used in our studies and/or to the combinatory effect of ligase and polymerase as it is in Gap-LCR. Direct comparison of ligase specificity in Gap-LCR to ligase specificity in the absence of polymerase in blunt-LCR is not feasible. Generation of target-independent ligation products is very common in blunt-LCR and would precede detection of products from mismatched targets. Moreover, the position of the mismatch in blunt-LCR probe and Gap-LCR probe cannot be directly compared; in Gap-LCR, depending on the size of the gap, a mismatch at the 3' end of the probe becomes a mismatch 2–4 bases away from the ligation junction after polymerase extension.

To determine whether discrimination obtained with Gap-LCR was solely due to the specificity of the polymerase or to the additive specificity of polymerase and ligase, we compared discrimination obtained with polymerase alone (in ASPCR) to discrimination obtained with polymerase plus ligase (in Gap-LCR) under the same reaction conditions. Our data indicate that the specificity of Gap-LCR depends on the fidelity of polymerase extension as well as on the specificity of ligation. The difference between Gap-LCR and ASPCR was further enhanced when the mismatch was positioned at the penultimate 3' base, since at that position the specificity increased for Gap-LCR but decreased for ASPCR when compared to the mismatch at the 3' end. Specificity of PCR could have been improved by optimizing conditions. However, these experiments were not aimed at comparing the performance of Gap-LCR versus ASPCR; they were designed to determine whether the specificity of Gap-LCR relied on polymerase alone or on polymerase plus ligase, thus the same reaction conditions needed to be utilized.

The potential of this amplification method to detect a mutated DNA sequence present at low copy number in a high background of wild-type DNA was evaluated. We observed that detection of mutant DNA in the presence of up to 10^4 -fold excess wild-type DNA was feasible. This result demonstrates the advantage of Gap-LCR over blunt-LCR. With blunt-LCR, the signal obtained from mutant DNA in the presence of 100-fold excess wild-type DNA could not be distinguished from the background noise (37). Our results suggest that Gap-LCR would allow detection of mutations present at frequencies as low as one in 10^4 gene copies;

thus Gap-LCR can be used in conditions where only a small fraction of cells are expected to contain the mutation.

In conclusion, we have demonstrated that Gap-LCR is a sensitive and specific amplification technique that can accurately discriminate single base changes. A significant advantage of the Gap-LCR assay described here is the ability to specifically detect the reaction products using a simple automated immunoassay system. Our study suggests that Gap-LCR can be used as a powerful tool in the diagnosis of genetic diseases, in monitoring drug resistant pathogens and in the detection of oncogenic mutations.

ACKNOWLEDGEMENTS

We would like to thank Dr Steve Wolinsky, Northwestern University, and Dr John Mellors, University of Pittsburgh, for their gift of HIV clones. We also thank our colleagues at Abbott Laboratories for their critical review of the manuscript.

REFERENCES

- 1 Kan, Y.W. and Dozy, A.M. (1978) *Lancet*, ii, 910-912.
- 2 Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. (1985) *Nature*, 313, 495-498.
- 3 Cotton, R.G.H., Rodrigues, N.R. and Campbell, R.D. (1988) *Proc. Natl. Acad. Sci. USA*, 85, 4397-4401.
- 4 Myers, R.M., Larin, Z. and Maniatis, T. (1985) *Science*, 230, 1242-1246.
- 5 Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 2766-2770.
- 6 Prosser, J. (1993) *Tibtech*, 11, 238-246.
- 7 Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 8278-8282.
- 8 Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986) *Nature*, 324, 163-166.
- 9 Gibbs, R.A., Nguyen, P.N. and Caskey, C.T. (1989) *Nucleic Acids Res.*, 17, 2437-2448.
- 10 Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Simmers, C., Kalsheker, N., Smith, J.C. and Markham, A.F. (1989) *Nucleic Acids Res.*, 17, 2503-2516.
- 11 Wu, D.Y., Ugozzoli, L., Pal, B.K. and Wallace, R.B. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 2757-2760.
- 12 Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) *Science*, 241, 1077-1080.
- 13 Wu, D.Y. and Wallace, R.B. (1989) *Genomics*, 4, 560-569.
- 14 Barány, F. (1991) *Proc. Natl. Acad. Sci. USA*, 88, 189-193.
- 15 Backman, K. (1989) European patent publication EP-A-0320 308 (priority filing 1987).
- 16 Laffler, T.G., Carrino, J.J. and Marshall, R.L. (1993) *Ann. Biol. Clin.*, 50, 821-826.
- 17 Bond, S., Carrino, J., Hample, H., Hanley, H., Reinhardt, L. and Laffler, T. *Papillomaviruses in Human Pathology. Recent Progress in Human Epidermoid Precancers*, Vol. 78. Raven Press, New York, NY, p. 425.
- 18 Dille, B.J., Butzen, C.C. and Birkenmeyer, L.G. (1993) *J. Clin. Microbiol.*, 31, 729-731.
- 19 Wu, D.Y. and Wallace, R.B. (1989) *Gene*, 76, 245-254.
- 20 Wiedmann, M., Wilson, W.J., Czajka, J., Luo, J., Barany, F. and Batt, C.A. (1994) *PCR Methods. Appl.*, 3S, 51-64.
- 21 Wiedmann, M., Czajka, J., Barany, F. and Batt, C.A. (1992) *Appl. Environ. Microbiol.*, 58, 3443-3447.
- 22 Wiedmann, M., Barany, F. and Batt, C.A. (1993) *Appl. Environ. Microbiol.*, 59, 2743-2745.
- 23 Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L. and Landegren, U. (1990) *Proc. Natl. Acad. Sci. USA*, 87, 8923-8927.
- 24 Backman, K., Carrino, J., Bond, S. and Laffler, T. (1991) European patent publication, EP-A-0439 182 (priority filing 1990).
- 25 Segev, D. (1990) PCT publication WO 90/01069 (priority filing 1988).
- 26 Birkenmeyer, L. and Armstrong, A.S. (1992) *J. Clin. Microbiol.*, 30, 3089-3094.
- 27 Birkenmeyer, L.G. and Mushahwar, I.K. (1991) *J. Virol. Methods*, 35, 117-126.
- 28 Hatt, C., Ward, M.E. and Clark, I.N. (1988) *Nucleic Acids Res.*, 16, 4053-4067.
- 29 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 30 Ugozzoli, L. and Wallace, R.B. (1991) *Methods: Companion Methods Enzymol.*, 2, 42-48.
- 31 Larder, B.A. and Kemp, S.D. (1989) *Science*, 246, 1155-1158.
- 32 Kwok, S., Kellogg, D.E., Spadic, D., Goda, L., Levenson, C. and Sninsky, J.J. (1990) *Nucleic Acids Res.*, 18, 999-1005.
- 33 Huang, M.M., Arnheim, N. and Goodman, M.F. (1992) *Nucleic Acids Res.*, 20, 4567-4573.
- 34 Cha, R.S., Zarbl, H., Keohavong, P. and Thilly, W.G. (1992) *PCR Methods Applic.*, 2, 14-20.
- 35 Bouma, S.R., Gordon, J., Hsieh, W.-T., Jou, T.-H., Beaudet, A.L. and Fang, P. (1994) PCT publication WO94/08047 (priority filing 1992).
- 36 Harada, K. and Orgel, L.E. (1993) *Nucleic Acids Res.*, 21, 2287-2291.
- 37 Kalin, I., Shepard, S. and Candrian, U. (1992) *Mutat. Res.*, 283, 119-123.

87. E. R. Fearon et al., *Science* 247, 49 (1990).
 88. K. W. Kinzler et al., *ibid.* 251, 1366 (1991).
 89. S. Srivastava, Z. Zou, K. Pirolo, W. Blattner, E. H. Chang, *Nature* 348, 747 (1990).
 90. S. Kwok and J. J. Sninsky, *PCR Technology: Principles and Applications for DNA Amplification*, H. Erlich, Ed. (Stockton, New York, 1989), pp. 235-244.
 91. T. J. White, R. Madec, D. H. Persing, *Adv. Clin. Chem.*, in press; S. D. Williams and S. Kwok, in *Laboratory Diagnosis of Viral Infections*, E. H. Lennette, Ed. (Dekker, New York, 1991), pp. 147-173.
 92. P. S. Walsh, D. A. Matzger, R. Higuchi, *Biotechniques* 10, 506 (1991).
 93. E. Blake, J. Mihalovich, R. Higuchi, P. S. Walsh, H. Erlich, unpublished data.
 94. J. Bill et al., *J. Exp. Med.* 169, 115 (1989); B. Budowle et al., *Am. J. Hum. Genet.* 48, 137 (1991).

95. R. C. Allen, G. Graves, J. Lowle, *Biotechniques* 7, 736 (1989); K. Kasai et al., *J. Forensic Sci.* 35, 1196 (1990).
 96. S. Scharf, unpublished data.
 97. R. Helmuth et al., *Am. J. Hum. Genet.* 47, 515 (1990).
 98. M. Stonckin, D. Hedgcock, R. Higuchi, L. Vigilant, H. A. Erlich, *ibid.* 48, 370 (1991).
 99. We thank our colleagues who contributed to the development and application of PCR. The space constraints of this review and the many publications on PCR prevent a comprehensive survey of advances and applications; we apologize to any of our colleagues whose studies have not been noted specifically. We are grateful to R. Saiki, S. Scharf, R. Higuchi, and R. Abramson for allowing us to cite their unpublished work; E. Rose for critical review; and K. Levenson for preparation of this manuscript.

(1) NOTICE: THIS MATERIAL MAY BE PROTECTED BY COPYRIGHT LAW (TITLE 17 U.S. CODE)

Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project

MARK D. ADAMS, JENNY M. KELLEY, JEANNINE D. GOCAYNE, MARK DUBNICK, MIHAEL H. POLYMERPOULOS, HONG XIAO, CARL R. MERRIL, ANDREW WU, BJORN OLDE, RUBEN F. MORENO, ANTHONY R. KERLAVAGE, W. RICHARD MCCOMBIE, J. CRAIG VENTER*

Automated partial DNA sequencing was conducted on more than 600 randomly selected human brain complementary DNA (cDNA) clones to generate expressed sequence tags (ESTs). ESTs have applications in the discovery of new human genes, mapping of the human genome, and identification of coding regions in genomic sequences. Of the sequences generated, 337 represent new genes, including 48 with significant similarity to genes from other organisms, such as a yeast RNA polymerase II subunit; *Drosophila* kinesin, *Notch*, and *Enhancer of split*; and a murine tyrosine kinase receptor. Forty-six ESTs were mapped to chromosomes after amplification by the polymerase chain reaction. This fast approach to cDNA characterization will facilitate the tagging of most human genes in a few years at a fraction of the cost of complete genomic sequencing, provide new genetic markers, and serve as a resource in diverse biological research fields.

THE HUMAN GENOME IS ESTIMATED TO CONSIST OF 50,000 to 100,000 genes, up to 30,000 of which may be expressed in the brain (1). However, GenBank lists the sequence of only a few thousand human genes and <200 human brain messenger RNAs (mRNAs) (2). Once dedicated human chromosome

sequencing begins in 5 years, it is expected that 12 to 15 years will be required to complete the sequence of the genome (3). It is therefore likely that the majority of human genes will remain unknown for at least the next decade. The merits of sequencing cDNA, reverse transcribed from mRNA, as a part of the human genome project have been vigorously debated since the idea of determining the complete nucleotide sequence of humans first surfaced. Proponents of cDNA sequencing have argued that because the coding sequences of genes represent the vast majority of the information content of the genome, but only 3% of the DNA, cDNA sequencing should take precedence over genomic sequencing (4). Proponents of genomic sequencing have argued the difficulty of finding every mRNA expressed in all tissues, cell types, and developmental stages and have pointed out that much valuable information from intronic and intergenic regions, including control and regulatory sequences, will be missed by cDNA sequencing (5). However, many genome enthusiasts have incorrectly stated that gene coding regions, and therefore mRNA sequences, are readily predictable from genomic sequences and have concluded that there is no need for large-scale cDNA sequencing. In fact, prediction of transcribed regions of human genomic sequence is currently feasible only for relatively large exons (6).

On the basis of our high output with automated DNA sequence analysis of 96 templates per day and consideration of the above issues, we initiated a pilot project to test the use of partial cDNA sequences (ESTs) in a comprehensive survey of expressed genes.

Sequence-tagged sites (STSs) are becoming standard markers for the physical mapping of the human genome (7). These short sequences from physically mapped clones represent uniquely identified map positions. ESTs can serve the same purpose as the random genomic DNA STSs and provide the additional feature of pointing directly to an expressed gene. An EST is simply a segment of a sequence from a cDNA clone that corresponds to an mRNA. ESTs longer than 150 bp were found to be the most useful for similarity searches and mapping.

M. D. Adams, J. M. Kelley, J. D. Gocayne, M. Dubnick, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. McCombie, and J. C. Venter are in the Section of Receptor Biochemistry and Molecular Biology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892. M. H. Polymeropoulos, H. Xiao, and C. R. Merrill are in the Laboratory of Biochemical Genetics, National Institute of Mental Health, Neuroscience Center at St. Elizabeth's Hospital, Washington, DC 20032.

*To whom correspondence should be addressed.

Libraries of Complementary DNAs

Of the estimated 30,000 genes expressed in the human brain, as many as 20,000 may encode low-abundance, brain-specific transcripts (1). The fact that up to one-fourth of all genetic diseases affect neurological functions is an indication of the diversity and importance of genes expressed in the brain (8).

An assumption in our choice of cDNA libraries was that random-primed and partial cDNA clones would be more informative in identifying genes and constructing a useful EST database than sequencing from the ends of full-length cDNAs (which contain 5' and 3' untranslated sequences) would be. By obtaining coding sequences, we hoped to take advantage of more sensitive peptide comparisons, in addition to nucleotide sequence comparisons. To discover the inherent limitations to be overcome in a large-scale cDNA sequencing project, we wanted to examine the diversity of representative cDNA libraries, identify desirable and undesirable characteristics of the libraries, and determine the information content and accuracy of single-run sequencing from both coding and flanking regions. Single-run sequencing involves performing a single sequence reaction, rather than relying on multiple, redundant reactions from each strand. We chose three commercial human brain cDNA libraries made from mRNA isolated from the hippocampus and temporal cortex of a 2-year-old female and from a fetal brain (9).

Single-run DNA sequence data were obtained from 609 randomly chosen cDNA clones (Table 1). Double-stranded cDNA clones in the pBluescript vector (Stratagene) were sequenced by a cycle sequencing protocol (10) with dye-labeled primers and 373A DNA Sequencers (Applied Biosystems). The average length of usable sequence was 397 bases with a standard deviation of 99 bases.

Subtractive hybridization has been used by researchers to reduce the population of highly represented sequences in a cDNA library (11, 12) by selectively removing sequences shared by another library. We tested subtractive hybridization as a way of enhancing the number of brain-specific clones in the hippocampus library by hybridizing the hippocampus library with a WI38 human lung fibroblast cell line cDNA library and removing the common sequences (Table 1) (12, 13).

Table 1. Composition of cDNA library determined by random clone sequencing. Each λ ZAP library (Stratagene) was converted en masse to Bluescript plasmids, transfected into *Escherichia coli* XL1-Blue (Stratagene) cells, and plated on plates with X-gal, isopropyl-1-thio- β -D-galactoside, and ampicillin. A total of 1058 clones were picked at random from three human brain cDNA libraries: fetal brain, 2-year-old hippocampus, and 2-year-old temporal cortex (9). Clones selected from the hippocampus library after subtraction with the fibroblast library are listed in the "Subtracted" column. Templates for DNA sequencing were PCR products or plasmids prepared by alkaline lysis method. About half of the templates prepared by PCR failed to yield an amplified fragment suitable for sequencing. This was primarily due to use of PCR conditions that minimized the need for further purification of the product but selected against amplification of long inserts (5 μ l of *coli* fresh or frozen overnight carrying the pBluescript plasmid, 7.5 μ M

EST Characterization

Initially, EST sequences were examined for similarities in the GenBank nucleic acid database (14). ESTs without exact GenBank

Table 2. EST matches to human genes. Matches of at least 97% were considered to indicate that the EST corresponds directly to the human gene. Number, GenBank accession number of the matched sequence. Map positions are from (8), except where indicated. LDL, low-density lipoprotein; EST names in GenBank are the three-digit number given here preceded by "EST00."

EST	Identification	Number	Map location
001	δ -Actin (cytoplasmic)	M10277	
002	δ -Actin (cytoplasmic)	M10277	
003	δ -Actin (cytoplasmic)	M10277	
264	γ -Actin (nonmuscle)	M24241	17p11-qter
265	γ -Actin (nonmuscle)	M24241	17p11-qter
005	CNPase	M19650	
006	CNPase	M19650	
237	ADP/ATP translocase	J03591	Xq13-q26
238	Fructose-1,6-bisphosphatase	X07292	17cen-q12
239	α -2-Macroglobulin	M11313	12p13.3-p12.3
240	α -Fodrin	M18627	9q33-34
242	α -Tubulin	K00558	†
243	α -Tubulin	K00558	†
004	δ -Tubulin	X02344	†
244	Amyloid A4	Y00264	21q21.3-22.05
245	Apolipoprotein J	J02908	8*
246	Breakpoint cluster region	X02596	22q11-q12
251	C-erbA- α -2	J03239	3p24.3
253	Catelectrin	J03578	
254	Calmodulin	J04046	
261	Elongation factor-1 α	X03558	†
262	Filaggrin	M24355	1q21
263	G β protein α subunit	X04408	20q13.2-q13.3
266	Glial fibrillary acidic protein	J04569	
268	Gln synthetase	Y00387	
280	Hexokinase	†	10p11.2
269	High-mobility group 1 protein	X12597	
278	LDL receptor-related protein	X13916	
284	Na ⁺ , K ⁺ -ATPase α subunit	X04297	1p13-p11
285	Neurofilament light chain	X05608	8p21
288	Phosphoglycerate kinase	L00160	Xq13
362	Ret proto-oncogene	M16029	10q11.2
363	RhoB	X05820	
366	Osteonectin	J03040	5q31-q33
367	Synaptophysin (p38)	X06389	Xp11.23-p11.22

*Indicates that the EST was mapped in this study by PCR. †The human hexokinase nucleotide sequence has been published (29) but does not appear in GenBank or EMBL. This EST was initially identified by matches to the mouse and rat nucleotide sequences and the human peptide sequence. ‡Mapping information on this isotype is not available.

each deoxynucleotide triphosphate, and 0.1 μ M each primer for 35 cycles: 94°C, 40 s; 55°C, 40 s; 72°C, 90 s). A further percentage of the PCR-generated templates failed to sequence, largely because of primer-dimer or other amplification artifacts. Qiagen columns (Studio City, California) improved the percentage of plasmid templates that yielded usable sequences from about 60% with a standard alkaline lysis protocol to over 90%. Overall, 117 PCR-generated templates and 497 plasmid templates gave usable sequences. Dideoxy chain-termination sequencing reactions were performed with fluorescent dye-labeled M13 universal or reverse primers (Applied Biosystems). After a cycle sequencing protocol (10), carried out in a Perkin-Elmer Thermal Cycler, sequencing reactions were run on a 373A automated DNA sequencer (Applied Biosystems). Some sequencing reactions were performed on an Applied Biosystems robotic workstation (28). For each column, numbers are indicated followed by percents in parentheses.

EST category	Hippocampus	Subtracted	Fetal brain	Temporal cortex
Database match—human				
Mitochondrial genes	48 (12.8)	10 (8.6)	3 (7.9)	6 (7.5)
Repeated sequences	39 (10.4)	14 (12.2)	6 (15.8)	0 (0)
Ribosomal RNA	10 (2.7)	7 (6.0)	0 (0)	11 (13.8)
Other nuclear genes	32 (8.6)	7 (6.0)	4 (10.5)	0 (0)
Database match—other	32 (8.6)	7 (6.0)	5 (13.2)	4 (5.0)
adenylate insert	160 (42.8)	44 (37.9)	20 (52.6)	6 (7.5)
insert	53 (14.1)	24 (20.7)	0 (0)	27 (33.7)
	1 (0.3)	3 (2.6)	0 (0)	26 (32.5)

matches were translated in all six reading frames, and each translation was compared with the protein sequence database Protein Information Resource (PIR) and the ProSite protein motif database

Table 3. EST similarities in the GenBank and PIR databases. All significant similarities ($P < 0.01$) with GenBank or PIR entries are listed. Matches indicate percent identical bases for nucleotides and percent similarity (identical plus conservative substitutions) for peptides. Number indicates the accession number or locus name of the matched sequence. Abbreviations used are as follows: B, bovine; BM, *Brugia malayi*; BMDV, bovine mucosal disease virus; C, chicken; CE, *Caenorhabditis elegans*; D, *Drosophila melanogaster*; E, *E. coli*; H, human; L, lamprey; M, mouse; N, *Neurospora crassa*; P, pig; PP, *Pseudomonas putida*; PRV, Pseudorabies virus; R, rat; S, squid; T, *Torpedo californica*; TN, transposon Tn 4556; X, *Xenopus laevis*; Y, yeast; UT, untranslated; MARCKS, myristoylated alanine-rich C kinase substrate; HPRT, hypoxanthine-guanine phosphoribosyltransferase; GTP, guanosine triphosphate; LAMP, lysosomal-associated membrane protein; tRNA, transfer RNA; snRNP, small nuclear ribonucleoprotein; IGF, insulin growth factor; Mito, mitochondrial; DBP, albumin promoter D site-binding protein; and Pol, polymerase. EST names in GenBank are the three-digit number given here preceded by "EST00."

EST	Description	Length	Match	Number
Nucleotide similarities (GenBank)				
247	80-87 kD MARCKS (B)	277	81.5	M24638
377	Mito ATPase δ subunit (B)	421	85.1	X06088
248	p ADP-ribosyltransferase substrate (B)	256	80	M27278
256	Enhancer of split (D)	264	71	M20571
257	Kinesin (D)	263	70.4	M24442
259	Xotch (X)	435	75.4	M33874
270	α -Tubulin (H)	495	82.3	X00734
271	α -Actinin (H)	272	85	X15804
273	Apollipoprotein A-I 5'-UT (H)	110	69	M20656
274	HPRT 3'-UT (H)	85	75	M26434
275	Kruppel-related Zn ²⁺ fingers (H)	88	67	M20678
276	LAMP-1 (H)	257	71.5	J04182
289	Aconitase (P)	318	89	J05224
293	ras-like (*)	71	74	X01669
295	IGF-binding protein 5'-UT (R)	115	77.3	J04486
299	ras-like (R)	138	57	X06889
300	RP L30 (R)	189	89	K02932
301	RP S10 (R)	273	90.8	X13549
365	UT conserved sequence element (H)	85	81	M24686
368	Electromotor neuron protein (T)	112	64	M30271
371	Maternal G10 mRNA (X)	234	80	X15243
372	Catalase T (Y)	65	72.3	X04625
374	RNA Pol II 6th subunit (Y)	216	64.7	M33924
Peptide similarities (PIR)				
247	80-87kD MARCKS (B)	62	82.3	S08341
377	Mito ATPase δ subunit (B)	97	92.8	S00763
249	GTP-binding protein smg p25A (B)	98	89.8	A35652
375	Genome polypeptide (BMDV)	27	74.1	GNWVBV
250	60K filarial antigen (BM)	109	78.0	A28209
252	Collagen 1 (CE)	57	57.9	A31219
255	Cadherin, neuronal (C)	42	64.3	A29964
256	Enhancer of split (D)	87	78.2	A30047
259	Notch (D)	102	72.5	A24768
260	Mobilization protein MbeA (E)	47	63.8	S04790
272	Ankyrin (H)	84	60.7	A35049
271	α -Actinin (H)	89	95.5	S05503
275	Finger protein XicGF20-1 (X)	30	80.0	S06565
279	Elongation factor Tu (*)	24	79.2	S06703
281	Monophenol monooxygenase (M)	29	69.0	YRMSC5
282	Neurogenic receptor trkB (M)	56	83.4	A35104
283	U1 snRNP 70K protein (M)	59	57.6	S04336
286	Leu-tRNA ligase (N)	48	58.3	A33475
287	Processing-enhancing protein (N)	97	79.4	S03968
289	Aconitase (P)	106	98.1	A35544
290	Pro-rich protein (clone cp7)	56	64.3	E25372
291	NtrA (PP)	31	61.3	JG0338
292	IE180 protein (PRV)	22	86.4	KDBEIT
293	ras-like (*)	53	58.5	B34788
294	Alcohol sulfotransferase (R)	35	71.4	A33569
296	Transcriptional activator DBP (R)	39	74.4	A34894
297	Myosin heavy chain (R)	60	58.3	MWRTS
298	Protein-tyrosine phosphatase (R)	22	86.4	A34845
299	ras-like (R)	55	58.2	TVHURR
300	RP L30 (R)	58	98.3	S11622
301	RP S10 (R)	67	97.0	S01881
364	Fibrinogen γ chain (L)	35	77.1	PGLMGS
257	Kinesin (S)	93	91.4	A35075
368	Electromotor neuron protein (T)	32	81.3	B33319
369	Hypothetical protein (TN)	37	64.9	JQ0431
370	Various actins (*)	37	75.7	S06062
371	Maternal G10 mRNA (X)	39	94.9	S05955
373	Hypothetical protein (Y)	24	75.0	C27061
374	RNA Pol II 6th subunit (Y)	73	90.4	B34588

*Matches with sequences from several organisms.

(14). Comparisons with the ProSite motif database were done by means of the program MacPattern from the EMBL Data Library (14a). GenBank and PIR searches were conducted with our modifications of the "basic local alignment search tool" programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (15). These modifications permit many query sequences to be automatically searched in a sequential fashion. PIR searches were run on the National Center for Biotechnology Information BLAST network service. The BLAST programs contain a rapid database-searching algorithm that searches for local areas of similarity between two sequences and then extends the alignments on the basis of defined match and mismatch criteria. The algorithm does not consider the potential of gaps to improve the alignment, thus sacrificing some sensitivity for 60- to 80-fold increase in speed over other database-searching programs such as FASTA (16).

Sequence similarities identified by the BLAST programs were considered statistically significant with a Poisson P -value < 0.01 . The Poisson P -value is the probability of as high a score occurring by chance, given the number of residues in the query sequence and the database. After the BLASTN search, 30 unmatched ESTs were compared against GenBank by FASTA to determine if significant matches were missed because of the use of BLASTN for the database search. No additional statistically significant matches were found. Statistical significance does not necessarily mean functional similarity; some of the matches reported here may indicate the presence of a conserved domain or motif or simply a common protein structure pattern. Statistically significant matches to GenBank and PIR are reported in Tables 2 and 3. The length and percent identity or similarity of each alignment is given in Table 3 to aid in evaluation of match quality.

On the basis of database searches, the 609 EST sequences were classified into eight groups as shown in Table 1. Four groups, with 197 of the sequences (32% of the total), consist of matches to human sequences: repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes. Forty-eight of the sequences (8%) matched nonhuman entries in GenBank or PIR, whereas 230 (38%) had no significant matches. The remaining 134 (22%) sequences contained no insert between the Eco RI cloning sites or consisted entirely of polyadenylate.

Table 4. Matches to the ProSite motif database. Pattern matches from the ProSite database (except posttranslational modification sites) are shown. Abbreviations used are as follows: AA, amino acyl; HIGH, motif consensus in single-letter amino acid code (30); ILGF, insulin-like growth factor; DHFR, dihydrofolate reductase; EGF, epidermal growth factor; Gal-P-UDP, galactose-1-phosphate-uridylyl; C2H2, two Cys and two His residues. EST names in GenBank are the three-digit number given here preceded by "EST00."

Motif name	EST
AA-tRNA ligase "HIGH"	094
ATP-binding site A	052,068,158,177,207,091,008,261
Carboxypeptidase/Zn ²⁺	112
COXI	249*
Cytochrome c	060,128,120,139,279*,218,063,106
DHFR	235
Elongation factor	261
EGF	187,203
2Fe/2S Ferredoxin	067
Gal-P-UDP-transferase	101
Glycoprotein hormone	112
ILGF-binding protein	193
Leu zipper	071,072,095,055,070,106,025,200,221,107,102,114,131,260*,290*,291*,294*,164,287*,061,369*
Nuclear localization	182,020,183,214,062
Rubredoxin	226
Snake toxin	085
Zinc finger (C2H2)	188,275*

*See Table 3 for ESTs with similarity to GenBank or PIR sequences.

Table 5. Accuracy of single-run double ESTs listed in Table 2 and those matched genes were aligned with sequences from GenBank with the GCG program BESTFIT. The first 85 nucleotides were the polylinker sequence that was not

aligned with the pBluescript SK reference sequence. Tabulation of errors began 15 bases into the BESTFIT alignment and thus is reported beginning with bases 101 to 200.

Bases from primer	Mismatches-ambiguities*	Gaps*		Accuracy %	Aligned bases
		Insertions	Deletions		
101-200	1.45				
201-300	1.72	0.18	0.19	98.2	8800
301-400	2.07	0.25	0.11	97.9	8130
>400	3.53	0.98	0.37	96.6	5404
		2.63	1.06	92.8	3197

*Error rates are reported as number of mismatches, insertions, or deletions per hundred aligned bases. "Mismatches" includes ambiguous base calls.

Thirty-six ESTs matched previously sequenced human nuclear genes with more than 97% identity (Table 2). Four of these ESTs were from genes encoding enzymes involved in maintaining metabolic energy, including ADP/ATP (adenosine diphosphate/adenosine triphosphate) translocase, aldolase C, hexokinase, and phosphoglycerate kinase. Human homologs of genes for the bovine mitochondrial ATP synthase F_0F_1 subunit and porcine aconitase were also found (Table 3). Brain-specific cDNAs included synaptophysin, glial fibrillary acidic protein (GFAP), and neurofilament light chain. At least six ESTs were from genes encoding proteins involved in signal transduction: 2',3'-cyclic nucleotide 3'-phosphodiesterase (CNPase) (two ESTs), calmodulin, *c-erbA*- α -2, G stimulating protein (G_s) α subunit, and Na^+ , K^+ -ATPase α subunit. Other ESTs were matches to genes for ubiquitous structural proteins—actins, tubulins, and fodrin (nonerythroid spectrin). Eight ESTs were from genes known to be associated with genetic disorders (8). More than half of the human-matched ESTs have been mapped to chromosomes, indicating the bias of GenBank entries toward well-studied genes and proteins.

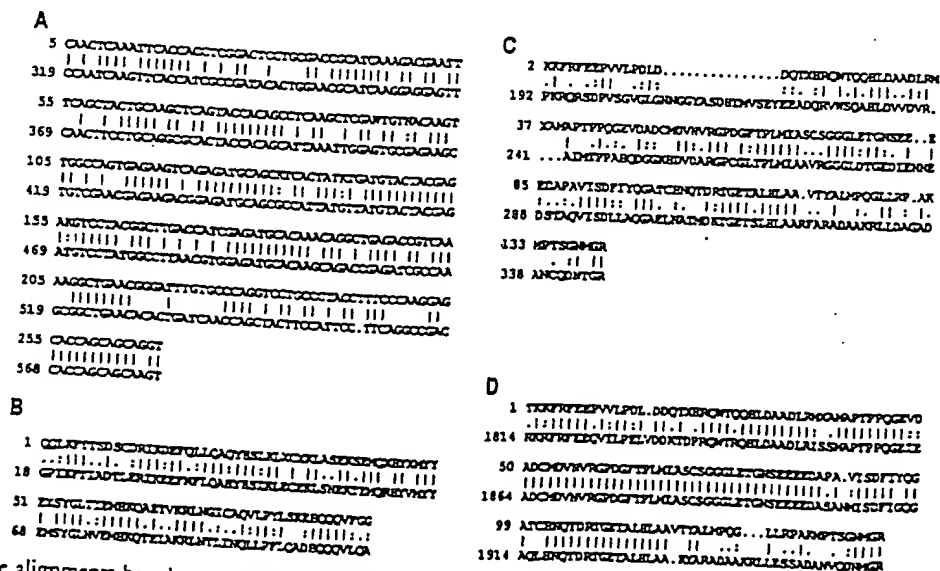
ESTs without significant GenBank matches were also compared to the ProSite database of recognized protein motifs. Not counting posttranslational modification signatures, 54 sequences contained motifs from the database (Table 4). Some patterns are found in scores or even, as in the case of the leucine zipper, hundreds of proteins that do not share the functional property implied by the presence of the motif.

Similarities to sequences from other organisms were also detected in the BLAST searches of GenBank and PIR (Table 3). Several ESTs were similar to "housekeeping" genes, including the ribosomal

proteins (RP) S10 and L30 (in the rat) and the above glycolytic enzymes. EST00257 showed strong nucleotide sequence similarity to the squid (67.4%) and *Drosophila* (70.4%) kinesin heavy chain. Kinesin was first described as a microtubule-associated motor protein involved in organelle transport in the squid giant axon (17). Six oncogene-related sequences were also among the cDNA clones sequenced. EST00299 and EST00283 showed similarity to several *ras*-related genes, and EST00248 matched the 3' untranslated region of the bovine substrate of botulinum toxin ADP-ribosyltransferase. We also observed similarities with a *Saccharomyces cerevisiae* RNA polymerase subunit and *Torpedo californica* electromotor neuron-associated protein. Two ESTs may represent new members of known human gene families: EST00270 matched the three β -tubulin genes with 88 to 91% identity and EST00271 matched α -actinin with 85% identity at the nucleotide level.

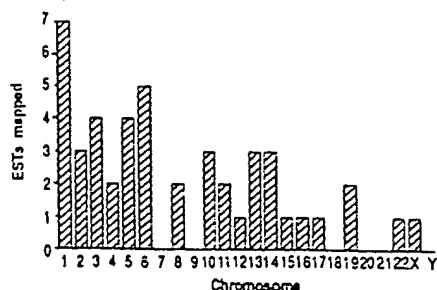
Among the most interesting of the primary sequence relationships was the similarity of ESTs to the *Drosophila* genes *Notch* and *Enhancer of split*. Nucleotide and peptide alignments of EST00256 and EST00259 with the *Drosophila* genes are shown in Fig. 1. Both genes are part of a signal cascade encoded by the "neurogenic" genes that are involved in the differentiation of neuronal and epidermal cell lineages in the neuroectoderm of the developing *Drosophila* embryo (18). It has been proposed that the *Enhancer of split* protein interacts with a membrane protein that is the product of the *Notch* gene to convert a developmental signal into an altered pattern of gene expression (18). EST00256 matched near the 5' end of the *Enhancer of split* coding sequence, away from the mammalian G protein β subunit and yeast *cdc4*-like elements (19). Part of the EST00259 match to *Notch* is in the *cdc10/SW16* region that is similar to three

Fig. 1. Sequence alignments of ESTs with *Drosophila* neurogenic genes. ESTs and EST translations were aligned with nucleotide and peptide sequences of two *Drosophila* neurogenic genes with the GCG program BESTFIT. The peptide alignment (30) of EST00259 with the *Xenopus* product, the *Xenopus laevis* homolog of *Notch*, is also shown. (A) EST00256 with *Drosophila Enhancer of split* (M20571); 69.202% identity; 1 gap. (B) EST00256 product with the product of *Drosophila Enhancer of split* (M20571); 72.826% similarity; 58.696% identity, 0 gaps. (C) EST00259 product with the *Drosophila Notch* product (K03508); 60.294% similarity; 43.382% identity; 5 gaps. (D) EST00259 product with the *Xenopus Notch* product (M33874); 32.143% similarity; 75.714% identity, 4 gaps. Gaps have been introduced to increase identity and similarity (indicated by dots in lines). Numbers in parentheses indicate the GenBank accession numbers. Symbols between lines: dashes indicate identity; double dots indicate a similarity score of 0.5 to 1.4; single dots represent a similarity score of 0.1 to 0.4. Scores are from pairwise alignments based on



the matrix of Schwartz and Dayhoff (31) as modified in the GCG package.

Fig. 2. Chromosome segregation of ESTs mapped by PCR. Chromosomes and ESTs are as follows: 1 (293*, 012, 077, 058, 079, 202, 086), 2 (021, 037, 234), 3 (248*, 257*, 274*, 062), 4 (009, 038), 5 (026, 030, 104, 123), 6 (301*, 007, 219, 023, 356), 8 (245*, 223), 10 (024, 197, 131), 11 (016, 111), 12 (014), 13 (372*, 273*, 200), 14 (221, 201, 008), 15 (165), 16 (373*), 17 (068), 19 (368*, 080), and X (276*). PCR conditions were as follows: 60 ng of genomic DNA was used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 μ Ci of α -³²P-labeled deoxycytidine triphosphate. The PCR was performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products were analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. Asterisks indicate those ESTs with similarity to GenBank or PIR sequences (Tables 2 and 3). EST names in GenBank are the three-digit number given here preceded by "EST00."



of the chromosome-specific material or the specific activity necessary for detection.

Automated DNA Sequencing Accuracy and GenBank Submission

ESTs that match human sequences in GenBank are excellent tools for the analysis of the accuracy of double-strand automated DNA sequencing. Ninety EST-GenBank matches were examined for the number of nucleotide mismatches and gaps required to achieve optimal alignment by the Genetics Computer Group (GCG) program BESTFIT (22). The number of mismatches, insertions, and deletions was counted for each hundred bases of the sequence (Table 5). As expected, the sequence quality was best closest to the primer and decreased rapidly after about 400 bases. The number of deletions and insertions relative to the GenBank reference sequence increased five- to tenfold beyond 400 bases, whereas the number of mismatches doubled. The average accuracy rate for individual double-stranded sequencing runs was 97.7% for up to 400 bases.

The minimum criteria for submission of ESTs to GenBank were that sequences be at least 150 bases in length and contain <3% ambiguous base calls. The overall accuracy of sequences submitted from each template group was at least 97%, based on matches to known human genes. Three hundred forty-eight ESTs met these criteria and were submitted to GenBank with accession numbers M61953 through M62300, inclusive. All ESTs except those matching mitochondrial or ribosomal RNA (rRNA) genes and simple repetitive elements were submitted to GenBank.

Conclusions and Prospects

Single-run DNA sequencing has proven to be an efficient method of obtaining preliminary data on cDNA clones. Our results demonstrate that sufficient information is contained in 150 to 400 bases of a nucleotide sequence from one sequencing run for preliminary identification of the cDNA and localization to a chromosome. In addition to the 35 ESTs homologous to known human genes, 48 ESTs matched sequences in GenBank or PIR with moderate to striking similarity, including high-quality matches with genes from such evolutionarily distant organisms as yeast (EST00374) and *Neurospora* (EST00287) (Table 3).

Two hundred thirty ESTs did not match any current database entries and therefore represent new, previously uncharacterized genes. A multitude of approaches for classifying these genes exists, including complete sequencing and expression, chromosome mapping, tissue distribution, and immunological characterization. Currently unidentified cDNAs will also be classified by similarity to genes from other organisms as those sequences become available. Three ESTs reported here (EST00257, EST00259, and EST0374) were identified by similarity to sequences that have appeared since the last full release of GenBank.

The random selection approach used here revealed an unacceptably large number of highly represented clones in these cDNA libraries. Over 30% of the clones from the hippocampus cDNA library consisted of rRNA, mitochondrial cDNAs, or inserts consisting entirely of polyA. Sixty-eight ESTs matched 12 different mitochondrial genes, including 18 matches to cytochrome oxidase I. Although elimination of these uninformative clones is a priority for developing ideal cDNA libraries, techniques to reduce repeated sequencing of clones will become increasingly important as large numbers of cDNAs are sequenced. The use of library preprocessing techniques such as subtraction, which preferentially reduces the

cell-cycle control genes in yeast and is tightly conserved in the *Xenopus laevis* *Notch* homolog, *Xotch*. In *Drosophila*, *Enhancer of split* is required for formation of epidermal tissue. *Notch* contains several epidermal growth factor-like repeats and appears to be involved in cell-cell communication during development (20).

Seven genes were represented by more than one EST. Comparisons of all the ESTs against one another revealed two overlaps of unknown ESTs: EST00233 and EST00234 matched in opposite orientations, and EST00235 and EST00236 matched in the same orientation beginning at the same nucleotide. Five human genes were represented by more than one EST: β -actin (three), γ -actin (two), α -tubulin (two), α -2-macroglobulin (two), and CNPase (two).

Mapping of ESTs to Human Chromosomes

We used the polymerase chain reaction (PCR) to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given EST (21). In this process, only the hybrids that contain the human gene corresponding to the EST will yield an amplified fragment. An EST is assigned to a chromosome by analysis of the segregation pattern of PCR products from hybrid DNA templates. The single human chromosome present in all hybrids that give rise to an amplified fragment is the location of the EST.

PCR mapping has been applied to 46 clones, as summarized in Fig. 2. The EST of the human gene for apolipoprotein J (also called SP-40,40 complement-associated protein and sulfated glycoprotein 2) was localized to chromosome 8. Eleven other ESTs with GenBank or PIR similarities were mapped to chromosomes. Although PCR mapping of somatic cell hybrids is relatively rapid—up to three clones can be assigned per day with a single thermal cycler—it is relatively expensive, costing about ten times as much as EST sequencing. With the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that have been proposed are multiplex in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific cDNA libraries. However, these methods are limited by the purity

population of certain sequences in the library (11, 12), and normalization, which results in all sequences being represented in approximately equal proportions in the library (23), should reduce repeated sequencing of high and intermediate abundance clones and maximize the chances of finding rare messages from specific cell populations. In our initial experiments with subtractive hybridization of the hippocampus library with a human fibroblast cDNA library, CNPase and GFAP clones were enriched greater than tenfold and twofold, respectively. Another characteristic of the ideal cDNA library would be directional cloning so that either a coding sequence or a 3' noncoding sequence could be selectively obtained.

The EST data, in conjunction with physical mapping, will provide a high resolution map of the location of genes along chromosomes, a map that would be more costly to construct by genomic sequencing and analysis. By performing a single DNA sequencing reaction on each cDNA clone, a key piece of information was obtained for the relatively low cost of about \$0.12 to \$0.15 per base. The EST approach will provide a new resource for the analysis of chromosome sequence and for human gene discovery.

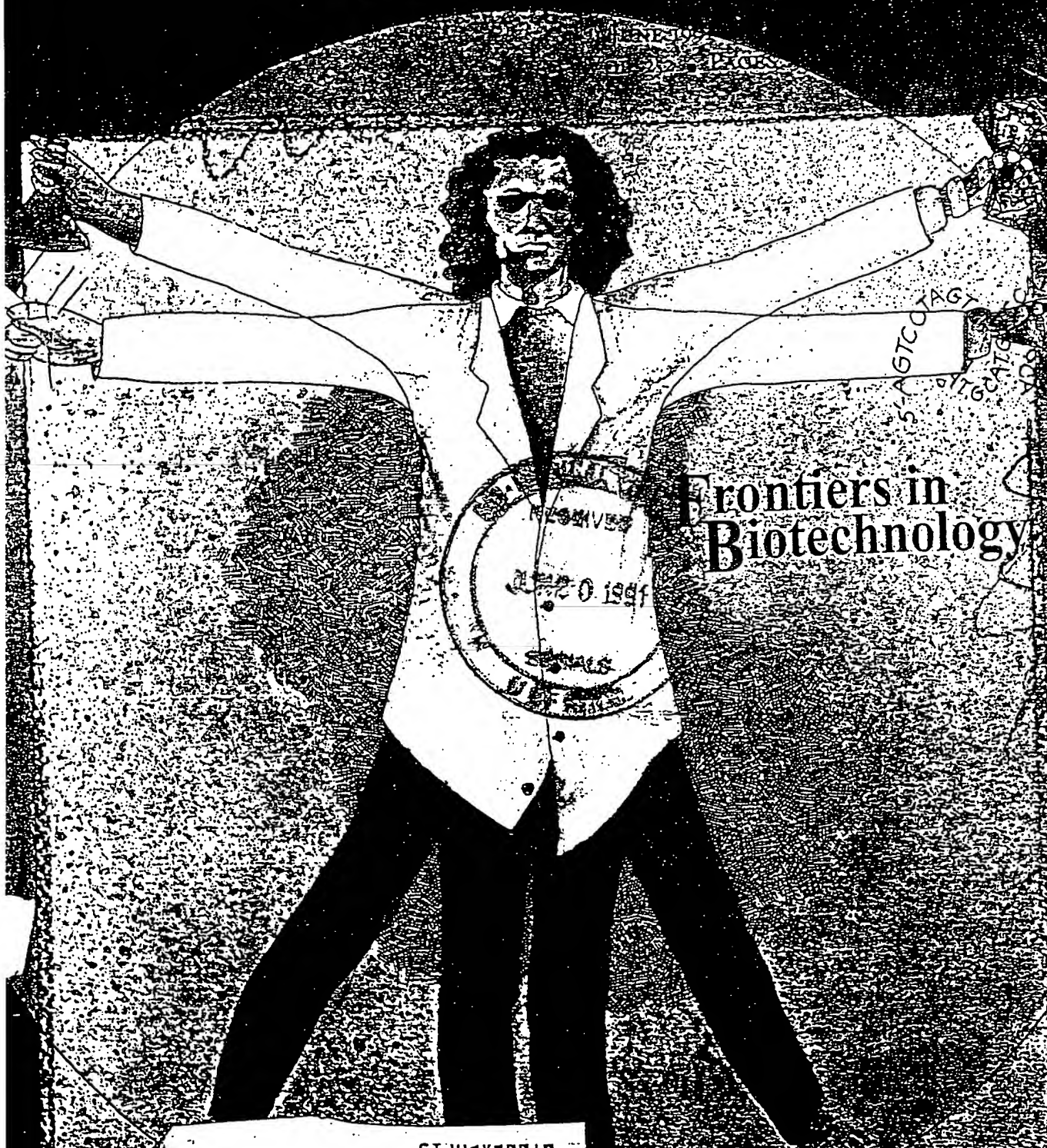
The screening of cDNA clones to identify the protein complement of a tissue has been explored by others to a limited extent. In 1983, Putney and co-workers sequenced over 150 clones from a rabbit muscle cDNA library and identified clones for 13 of the 19 known muscle proteins, including one new isotype, but no unknown coding sequences (24). Over 400 adult head-specific cDNA clones from *Drosophila* have been identified by differential screening of cDNA libraries from different developmental stages (25). Improvements in DNA sequencing technologies have now made feasible essentially complete screening of the expressed gene complement of an organism.

In our own laboratory, the EST approach should result in the partial sequencing of most human brain cDNAs in a few years. Similar approaches begun elsewhere (26) could result in a database of most human expressed genes in less than 5 years. The presence of these minimally characterized sequences in GenBank will assist research efforts in several areas of biology. The EST database will provide identification and confirmation of coding regions in naive genomic sequences. Sublocalization of cDNAs that have been mapped to chromosomes will help define the genetic content of specific chromosomal regions and permit correlation with patterns of inheritance in genetic disease. In a related experiment, chromosome sublocalization was the key to establishing that the γ -aminobutyric acid-benzodiazepine receptor β_3 subunit is deleted in individuals with Angelman-Prader-Willi syndrome (27). We anticipate that ESTs from human brain will further the identification of genes associated with other neurological diseases and will provide a more complete view of gene expression in the brain.

REFERENCES AND NOTES

1. J. G. Sutcliffe, *Annu. Rev. Neurosci.* 11, 157 (1988).
2. GenBank Release 66.0, December 1990.
3. D. Baltimore, Ed., "Report of the Ad Hoc Program Advisory Committee on Complex Genomes," Reston, VA, February 1988 (National Institutes of Health, Bethesda, MD, 1988).
4. S. Brenner, *Ciba Found. Symp.* 149, 6 (1990).
5. "Report of the Committee on Mapping and Sequencing the Human Genome" (National Academy Press, Washington, DC, 1988).
6. J. Fickett, *Nucleic Acids Res.* 10, 5303 (1982).
7. M. Olson, L. Hood, C. Cantor, D. Botstein, *Science* 245, 1434 (1989).
8. V. A. McKusick, *Online Mendelian Inheritance in Man*, from the Welch Medical Library, Johns Hopkins University School of Medicine, Baltimore, MD.
9. Stratagene catalog numbers 936206, 936205, and 935205, respectively.
10. Cycle sequencing was performed in a Perkin-Elmer Thermal Cycler for 15 cycles of 95°C, 30 s; 60°C, 1 s; 70°C, 60 s; and 15 cycles of 95°C, 30 s; 70°C, 60 s with the Taq Dye Primer Cycle Sequencing Core Kit protocol (Applied Biosystems).
11. D. W. Schmid and C. Girou, *J. Neurochem.* 48, 307 (1987); J. Fargnoli, N. J. Holbrook, A. J. Fornace, Jr., *Anal. Biochem.* 187, 364 (1990); J. R. Duguid and M. C. Dinauer, *Nucleic Acids Res.* 18, 2789 (1990); F. H. Travis and J. G. Sutcliffe, *Proc. Natl. Acad. Sci. U.S.A.* 85, 1696 (1988); K. Kato, *Eur. J. Neurosci.* 2, 704 (1990).
12. C. W. Schweinfest et al., *Genet. Anal. Tech. Appl.* 7, 64 (1990).
13. H. L. Sive and T. St. John, *Nucleic Acids Res.* 16, 10937 (1988).
14. GenBank Release 65.0 and collective updates through 21 January 1990, PIR Release 26.0, and ProSite (EMBL protein motif database) Release 5.0 were used.
15. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. Lipman, *J. Mol. Biol.* 215, 403 (1990).
16. W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85, 2444 (1988).
17. R. D. Vale, T. S. Reese, M. P. Sheetz, *Cell* 42, 39 (1985).
18. J. A. Campos-Ortega, *Trends Neurosci.* 11, 400 (1988).
19. D. A. Hartley, A. Preiss, S. Artavanis-Tsakonas, *Cell* 55, 785 (1988); C. Klämbt, E. Knust, K. Tietze, J. A. Campos-Ortega, *EMBO J.* 8, 203 (1989).
20. U. Banerjee and S. L. Zipursky, *Neuron* 4, 177 (1990).
21. We designed oligonucleotide primer pairs from EST sequences to minimize the chance of amplifying through an intron. Introns are less likely to be present in 3' untranslated sequences, so ESTs were examined for the presence of stop codons in each reading frame and for consensus splice junctions [J. D. Hawkins, *Nucleic Acids Res.* 16, 9893 (1988); M. B. Shapiro and P. Senapathy, *ibid.* 15, 7155 (1987)]. Also desirable for mapping are 3' untranslated sequences because they are less likely to be identical in other members of a gene family or in pseudogenes. We used the primers in a PCR to amplify from total human genomic DNA templates as described in the legend to Fig. 2. If the resulting product was of the expected size, then the PCR reaction was repeated with DNA templates from two panels of human-rodent somatic cell hybrids: Bios PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).
22. J. Devereux, P. Haeberli, O. Smithies, *Nucleic Acids Res.* 12, 387 (1984).
23. S. R. Patanjali, S. Parimoo, S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* 88, 1943 (1991); M. S. H. Koch, *Nucleic Acids Res.* 18, 5705 (1990).
24. S. D. Putney, W. C. Herlihy, P. Schimmel, *Nature* 302, 718 (1983).
25. M. J. Palazzolo et al., *Neuron* 3, 527 (1989).
26. *Hum. Genome News* 2 (no. 6), 1 (1991).
27. J. Wagstaff et al., *Am. J. Hum. Genet.*, in press.
28. R. Cathcart, *Nature* 347, 310 (1990).
29. S. Nishi, S. Seino, G. I. Bell, *Biochem. Biophys. Res. Commun.* 157, 937 (1988).
30. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
31. R. O. Schwartz and M. O. Dayhoff, in *Atlas of Protein Sequences and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1979), pp. 353-358.
32. Dedicated to the memory of J. E. Venter, 18 August 1923 to 10 June 1982.

SCIENCE



***** CAR-RT SORT ** CR55
47 000018948725C 12/20/91N9125
UNIVERSITY MINNESOTA
BIOMEDICAL LAB 325A DIEHL
505 ESSEX ST
MINNEAPOLIS MN 5545

Genetic disease detection and DNA amplification using cloned thermostable ligase

(β -globin gene/ligase chain reaction/sickle-cell allele/single-base mutation)

FRANCIS BARANY

Department of Microbiology, Hearst Microbiology Research Center, Cornell University Medical College, 1300 York Avenue, New York, NY 10021

Communicated by Hamilton O. Smith, September 26, 1990 (received for review July 25, 1990)

ABSTRACT Polymerase chain reaction, using thermostable DNA polymerase, has revolutionized DNA diagnostics. Another thermostable enzyme, DNA ligase, is harnessed in the assay reported here that both amplifies DNA and discriminates a single-base substitution. This cloned enzyme specifically links two adjacent oligonucleotides when hybridized at 65°C to a complementary target only when the nucleotides are perfectly base-paired at the junction. Oligonucleotide products are exponentially amplified by thermal cycling of the ligation reaction in the presence of a second set of adjacent oligonucleotides, complementary to the first set and the target. A single-base mismatch prevents ligation/amplification and is thus distinguished. This method was exploited to detect 200 target molecules as well as to discriminate between normal β^A - and sickle β^S -globin genotypes from 10- μ l blood samples.

DNA diagnostics uses the tools of molecular biology to identify nucleotide substitutions, deletions, or insertions in genes of medical interest (1). A reliable DNA diagnostics method will require faithful amplification of target sequences, accurate single-base discrimination, low background, and, ultimately, complete automation. The initial target nucleic acid amplification may be accomplished by using the polymerase chain reaction (PCR) (2), self-sustained sequence replication (3), or ligase amplification reaction (4, 5). Subsequently, single-base mismatches may be detected via allele-specific and reverse oligonucleotide hybridization (6, 7), denaturing gradient gel electrophoresis (8), RNase or chemical cleavage of mismatched heteroduplexes (9, 10), use of nucleotide analogs (11), or fluorescence PCR amplification/detection (12).

Landegren *et al.* (13) have pioneered an oligonucleotide ligation assay to circumvent the need for electrophoresis or precise hybridization conditions. Two oligonucleotide probes are hybridized to denatured DNA, such that the 3' end of the first one is immediately adjacent to the 5' end of the second probe. DNA ligase can covalently link these two oligonucleotides, provided that the nucleotides at the junction are perfectly base-paired to the target (4, 5, 13, 14). A single-nucleotide substitution can, therefore, be distinguished. Use of biotin on the first probe and a suitable nonisotopic reporter group on the second probe allows for product capture and detection (13) in a manner amenable to automation.

Ideally, the oligonucleotides should be sufficiently long (20–25 nucleotides) so that each will preferentially hybridize to its unique position on the human genome. The specificity of ligation should be particularly enhanced by performing the reaction at or near the melting temperature (t_m) of the two oligonucleotides. At higher temperatures a single-base mismatch at the junction forms not only an imperfect double

helix but also destabilizes hybridization of the mismatched oligonucleotide.

This report describes DNA detection that uses a thermostable ligase to exquisitely discriminate between a mismatched and complementary DNA helix (Fig. 1 *Upper*). Because the enzyme retains activity after multiple thermal cycles, the ligations may be repeated to linearly increase product [termed ligase detection reaction (LDR)]. Product may be further amplified in a ligase chain reaction (LCR) by using both strands of genomic DNA as targets for oligonucleotide hybridization. Two sets of adjacent oligonucleotides, complementary to each target strand, are used. The ligation products from one round can become the targets for the next round of ligation (Fig. 1 *Upper*). By use of LCR, the amount of product can be increased in an exponential fashion by repeated thermal cycling.

MATERIALS AND METHODS

Thermostable Ligase. Plasmid libraries of *Thermus aquaticus* strain HB8 DNA (ATCC27634) were screened for the ability to complement a temperature-sensitive *ligts7* derivative of *Escherichia coli* [unpublished work; ref. 16]. One complementing plasmid (pDZ1) contained a thermostable ligase gene as evidenced by (i) presence of a thermostable NAD⁺-dependent nick-closing (ligase) activity in crude extracts when assayed at 65°C (17) and (ii) DNA sequence analysis of the first 60 codons of the putative gene revealed >50% amino acid identity to *E. coli* ligase (18). Thermostable ligase was purified from *E. coli* cells containing the ligase gene cloned downstream of an inducible T7 expression system (19), as described elsewhere (unpublished work). Ligase activity was assayed for the ability to seal nicked plasmid DNA (pUC4K1XX) as monitored by electrophoresis on 1% agarose gel. One nick-closing unit of ligase is defined as the amount of ligase that circularizes 0.5 μ g of nicked pUC4K1XX DNA in 20 μ l of 20 mM Tris-HCl, pH 7.6/50 mM KCl/10 mM MgCl₂/1 mM EDTA/10 mM NAD⁺/10 mM dithiothreitol overlaid with a drop of mineral oil after 15-min incubation at 65°C.

Genomic DNA, Plasmid DNA, and Oligonucleotides. Human genomic DNA was isolated from 0.5 ml of whole blood as described (20). Proteinase K and RNase A were removed by sequential extractions with phenol, phenol/chloroform, chloroform, 1-butanol (twice), and nucleic acid was recovered by precipitation with ethanol. Samples were boiled for 5 min before use in LCR assays. Plasmid DNAs containing the β^A - and β^S -globin gene alleles were a gift from D. Nickerson (California Institute of Technology, Pasadena, CA) and were digested with *Taq* I before use as target DNA. Oligonucleotides were assembled by the phosphoramidite method (21) on an Applied Biosystems model 380A DNA synthesizer, purified by reversed-phase HPLC, and provided

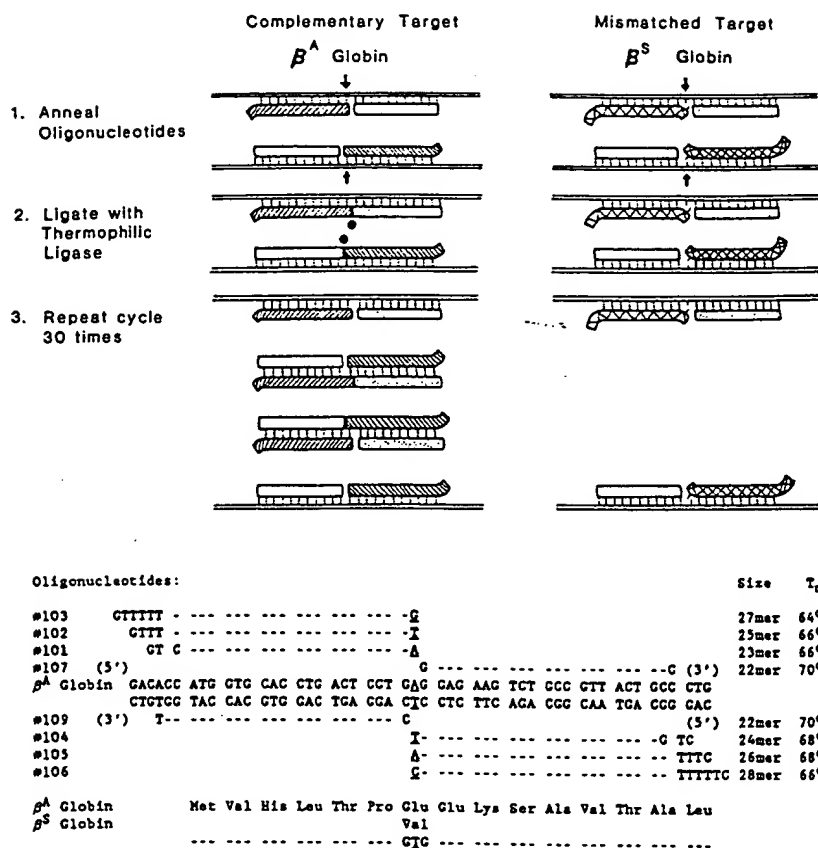


FIG. 1. (Upper) Diagram depicting DNA amplification/detection by using LCR. DNA is heat denatured, and four complementary oligonucleotides are hybridized to the target at a temperature near their melting temperature (65°C; t_m). Thermostable ligase will covalently attach only adjacent oligonucleotides that are perfectly complementary to the target (Left). Products from one round of ligations become targets for the next round, and thus products increase exponentially. Oligonucleotides containing a single-base mismatch at the junction do not ligate efficiently and, therefore, do not amplify product (Right). (Lower) Nucleotide sequence and corresponding translated sequence of the oligonucleotides used in detecting β^A - and β^S -globin genes. Oligonucleotides 101 and 104 detect the β^A target, whereas oligonucleotides 102 and 105 detect the β^S target when ligated to labeled oligonucleotides 107 and 109, respectively. Oligonucleotides 103 and 106 were designed to assay the efficiency of ligation of G-T or G-A and C-A or C-T mismatches when using β^A - or β^S -globin gene targets, respectively. Oligonucleotides have calculated t_m values of 66–70°C (15), just at or slightly above ligation temperature. The diagnostic oligonucleotides (101–106) contained slightly different length tails to facilitate discrimination of various products when separated on polyacrylamide denaturing gel.

by R. Kaiser and S. Horvath (California Institute of Technology, Pasadena, CA). Oligonucleotide sequences (5'-3') are: 101, GTCATGGTGCACCTGACTCCTGA; 102, GTTTCATGGTGCACCTGACTCCTGT; 103, GTTTTTCATG-GTGCACCTGACTCCTGG; 104, CTGCAGTAACGGCAGACTTCTCTCT; 105, CTTTGCAGTAACGGCAGACTTCTTCCA; 106, CTTTTCGAGTAACGGCAGACTTCTCTCC; 107, GGAGAAGTCTGCCGTTACTGCC; 109, CAGGAGT-CAGGTGCACCATGGT. (See Fig. 1.)

³²P Labeling of Oligonucleotides. Oligonucleotides 107 or 109 (0.1 μ g = 15 pmol) were 5' end-labeled in 20 μ l of 30 mM Tris-HCl, pH 8.0/20 mM Tricine/10 mM MgCl₂/0.5 mM EDTA/5 mM dithiothreitol/400 μ Ci of [γ -³²P]ATP (6,000 Ci/mM = 60 pmol ATP, New England Nuclear; 1 Ci = 37 GBq) by addition of 15 units of T4 polynucleotide kinase (New England Biolabs). After incubation at 37°C for 45 min, unlabeled ATP was added to 1 mM, and incubation was continued an additional 2 min at 37°C. The reaction was terminated by adding 0.5 μ l of 0.5 M EDTA, and the kinase was heat-inactivated (65°C for 10 min). Unincorporated ³²P label was removed by chromatography with Sephadex G-25 pre-equilibrated with Tris/EDTA buffer. Specific activity ranged from 7 to 10 \times 10⁸ cpm/ μ g of oligonucleotide.

LDR and LCR Reaction Conditions. For LDR reactions, labeled oligonucleotide (200,000 cpm = 0.28 ng = 40 fmol) and unlabeled diagnostic oligonucleotide (0.27 ng = 40 fmol) were incubated in the presence of target DNA (1 fmol = 6 \times 10⁸ molecules of *Taq* I-digested β^A - or β^S -globin plasmid) in 10 μ l of 20 mM Tris-HCl, pH 7.6/100 mM KCl/10 mM MgCl₂/1 mM EDTA/10 mM NAD⁺/10 mM dithiothreitol/4 μ g of salmon sperm DNA/15 nick-closing units of *T. aquaticus* ligase and overlaid with a drop of mineral oil. Reactions were incubated at 94°C for 1 min followed by 65°C for 4 min, and this cycle was repeated 5 or 20 times. For LCR reactions, unlabeled diagnostic oligonucleotide pairs (101 and 104, 102 and 105, or 103 and 106; 40 fmol each) and adjacent pairs of labeled oligonucleotides (107 and 109, 40 fmol each) were

incubated in the presence of ligase and target DNA (ranging from 100 amol to less than one molecule per tube) with 20 or 30 cycles as described above.

Electrophoresis. Samples (4 μ l) were in 45% formamide and denatured by boiling for 3 min before loading (40,000 or 80,000 cpm/lane). Electrophoresis was in 10% polyacrylamide gel containing 7 M urea in a buffer of 100 mM Tris borate, pH 8.9/1 mM EDTA for 2 hr at 60-W constant power. After removing urea, gels were dried and autoradiographed overnight at -70°C on Kodak XAR-5 film with the aid of a Cronex intensifying screen (DuPont).

RESULTS

The gene encoding human β -globin was selected as a model system to test ligation amplification and detection. The normal β^A and sickle β^S genes differ by a single A \rightarrow T transversion that leads to a change of a glutamic acid residue to a valine in the hemoglobin β chain [Fig. 1, Lower (22)]. Diagnostic oligonucleotides containing the 3' nucleotide unique to each allele were synthesized with different-length 5' tails (Fig. 1 Lower). Upon ligation to the invariant ³²P-labeled adjacent oligonucleotide, the individual products could be distinguished when separated on a polyacrylamide denaturing gel and detected by autoradiography.

Specificity of Thermostable Ligase. The specificity of ligating oligonucleotide pairs on a target DNA with perfect complementarity was directly compared with each possible mismatch (see Fig. 2 and Table 1). Results show that *T. aquaticus* ligase efficiently links correctly base-paired oligonucleotides and gives near zero ligation in the presence of a mismatch (Table 1). When only 1 fmol of target DNA was used under LDR conditions, the worst mismatches were 1.5–1% (G-T, T-T), whereas other mismatches were <0.4% (A-A, C-T, G-A, G-A) of the products formed with complementary oligonucleotide base pairs (A-T). This is substan-

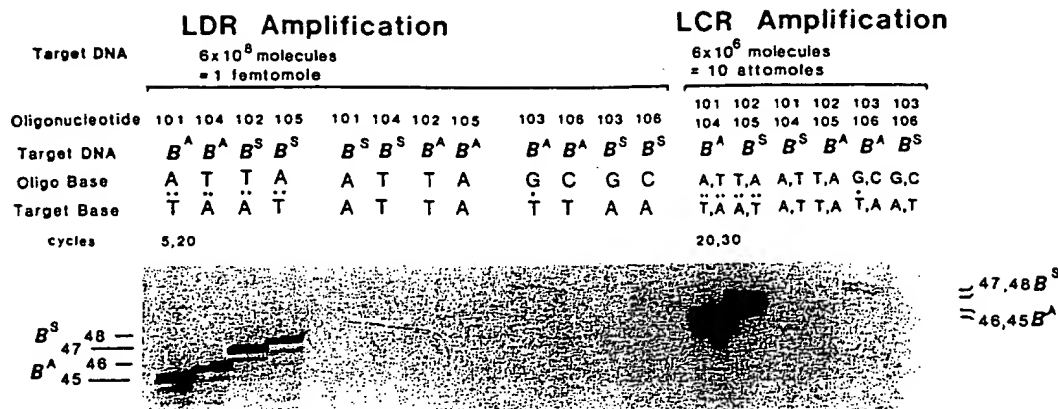


FIG. 2. Autoradiogram showing specificity of *T. aquaticus* ligase under LDR and LCR amplification conditions. Specificity was assayed by ligation of diagnostic oligonucleotides in the presence of either complementary or mismatched β^A - or β^S -globin gene target DNA (LDR amplification). Ligation of diagnostic oligonucleotides 101 (β^A allele), 102 (β^S allele), or 103 to labeled 107 gives lengths of 45, 47, or 49 nucleotides, respectively. For the complementary strand, ligation of diagnostic oligonucleotides 104 (β^A allele), 105 (β^S allele), or 106 to labeled 109 gives lengths of 46, 48, or 50 nucleotides, respectively. The diagnostic oligonucleotide listed in each lane and the appropriate adjacent labeled oligonucleotide (40 fmol each) was incubated with target DNA (1 fmol = 6×10^8 molecules of *Taq* I-digested β^A - or β^S -globin plasmid), as described. In LCR amplification, samples contained pairs of diagnostic oligonucleotides (β^A allele-specific 101 and 104, β^S allele-specific 102 and 105, or "C-G pair" 103 and 106), both labeled oligonucleotides (107 and 109), and were incubated with ligase and 10 amol of target DNA (6×10^6 molecules; 100-fold less than for LDR) as described. Samples were loaded in groups of eight and run into the gel; then the next set was loaded. This accounts for the "slower" migration of bands on the right side of the autoradiogram. (Intensifying screen was not used for this autoradiogram.) Bands were excised from the gel and assayed for radioactivity (Table 1).

tially better than found for mesophilic T4 or *E. coli* ligase when using similar radioactive detection methods (13, 14).

In the amplification/detection (LCR) experiments, four oligonucleotides were incubated with ligase and 10 amol of target DNA (see Fig. 2 Right and Table 1 lower part). The 3' nucleotide of each unlabeled diagnostic oligonucleotide was either complementary or mismatched to the target DNA and yet was always complementary to its pair—i.e., A-T for 101 and 104, T-A for 102 and 105, and G-C for 103 and 106.

Table 1. Quantitation of complementary and mismatched LDR and LCR

Amplification	Oligonucleotide base-target base	Product formed, %*	Mismatched/complementary, %†
LDR (6×10^8 target molecules = 1 fmol)	A-T	21.5	
	T-A	13.2	
	T-A	17.9	
	A-T	12.4	
	A-A	<0.1	<0.4
	T-T	0.12	0.7
	T-T	0.16	1.0
	A-A	<0.1	<0.4
	G-T	0.30	1.4
	C-T	<0.1	<0.4
LCR (6×10^6 target molecules = 10 amol)	G-A	<0.1	<0.4
	C-A	<0.1	<0.4
	A-T, T-A	41.4	
	T-A, A-T	10.4	
	A-A, T-T	0.45	1.1
	T-T, A-A	<0.05	<0.2
	G-T, C-A	0.51	1.3
	G-A, C-T	<0.05	<0.2

Bands from 20-cycle LDR and 30-cycle LCR experiments described in Fig. 2 were excised from the gels and assayed for radioactivity.

*Percentage product formed = cpm in product band/cpm in starting oligonucleotide band.

†Percentage mismatched/complementary = cpm in band of mismatched oligonucleotide/cpm in band of complementary oligonucleotide when using the same target DNA and indicates noise-to-signal ratio.

Four-way (target independent) ligation was minimized by use of (i) carrier salmon sperm DNA and (ii) oligonucleotides designed to create single-base 3' overhangs (this work, see Fig. 1) or single-base 5' overhangs (not tested). Note that an initial "incorrect" ligation of a mismatched oligonucleotide to target DNA would subsequently be amplified with the same efficiency as a correct ligation (See Fig. 1). Nevertheless, the worst mismatches were 1.3% to 0.6% (G-T, C-A; A-A, T-T), whereas others were <0.2% (T-T, A-A; G-A, C-T) of the products formed with complementary basepairs (A-T, T-A). LCR, using thermostable ligase, is thus the only method that can both amplify and detect single-base mismatches with high signal-to-noise ratios (4, 5).

The entire set of experiments described above was repeated with a buffer containing 150 mM instead of 100 mM KCl. Results were essentially the same as in Fig. 2 and Table 1; mismatches for LDR ranged from 0.6% to <0.3% and for LCR ranged from 1.7% to <0.3% of the complementary products (data not shown). Thus for *T. aquaticus* ligase, discrimination between matched and mismatched oligonucleotides is not critically dependent on salt conditions, in contrast to the requirements for mesophilic ligases (4, 5, 13, 14).

Specificity of LCR DNA Amplification with Sub-amol Quantities of Target DNA. The extent of LCR DNA amplification was determined in the presence of target DNA ranging from 100 amol = 6×10^7 molecules to <1 molecule per tube (Fig. 3, Table 2). In the absence of target DNA, no background signal was detected when carrier salmon sperm DNA (4 μ g) was present (compare last 8 lanes of Fig. 3). At higher target concentration, DNA amplification was essentially complete after 20 cycles, whereas at lower initial target concentration substantially more product is formed with additional amplification cycles. After 30 cycles of LCR, 200 molecules of initial target DNA were amplified 1.7×10^5 fold and thus could be readily detected. The average efficiency of ligation per cycle (40–50%, calculated as described in ref. 4) could be potentially enhanced by altering buffer conditions [such as using NH_4Cl , MnCl_2 , polyamines, or polyethylene glycols (17)], enzyme concentration, or thermal-cycling times and temperatures.

LCR Amplification

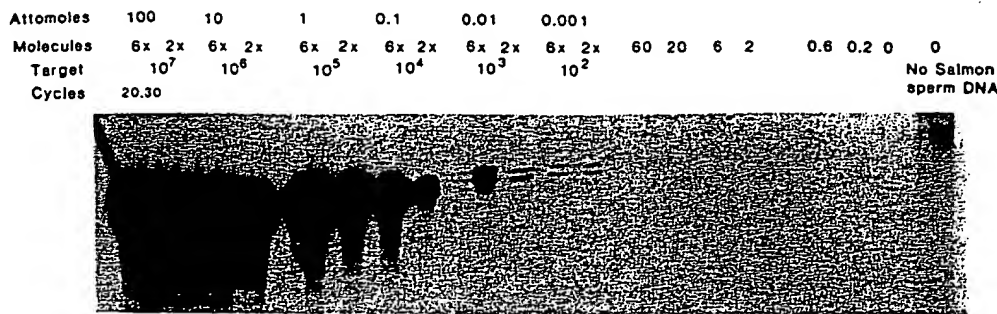


FIG. 3. Autoradiogram showing LCR amplification at different target concentrations. Labeled invariant oligonucleotides (107 and 109; 200,000 cpm = 40 fmol each) and unlabeled β^A allele oligonucleotides (101 and 104; 40 fmol each) were incubated with target DNA (ranging from 100 amol = 6×10^7 molecules to <1 molecule per tube of *Taq* I-digested β^A -globin plasmid) and ligase as described. Samples were electrophoresed, gel was autoradiographed overnight, and bands were counted as described (see Table 2). Bands of 45 and 46 nucleotides correspond to ligation products of the coding and complementary β^A -globin oligonucleotides. Lower-molecular-mass products correspond to ligation of minor species in the synthesized oligonucleotide preparations that were shorter than intended product. Samples were loaded in groups of eight, giving the appearance of slower migration on the right of the autoradiogram.

To test ligase discrimination between complementary and mismatched oligonucleotides in a direct competition assay, the above LCR experiment was repeated with or without oligonucleotides that would give G-T and C-A mismatches (see Table 3). At higher target concentrations, the mismatched product ranged from 1.8% to 0.5% of the complementary product. Mismatched product could not be detected when using <3 amol of target DNA. As control, excess mismatched target DNA (β^S instead of β^A -globin DNA at 6×10^7 molecules per tube) gave only 2.1% and 1.5% product. Thus, the signal from the correctly paired ligation products is 50- to 500-fold higher than from mismatched products, under either competition or individual LCR ligation conditions.

Detection of β -Globin Alleles in Human Genomic DNA. DNA isolated from the blood of normal ($\beta^A\beta^A$), carrier ($\beta^A\beta^S$), and sickle cell ($\beta^S\beta^S$) individuals was tested for allele-specific LCR detection. With target DNA corresponding to 10 μ l of blood, β^A and β^S alleles could be readily

detected by using allele-specific LCR (Fig. 4). As seen with plasmid-derived target DNA (see Fig. 2), efficiency of ligation (and hence detection) is somewhat less for β^S - than β^A -specific oligonucleotides. This difference may be a function of the exact nucleotide sequence at the ligation junction or the particular oligonucleotides (with differing 5' tails) used in these LCR experiments. Nevertheless, the results show the feasibility of direct LCR allelic detection from blood samples without any need for primary PCR or self-sustained sequence replication amplification.

DISCUSSION

The specificity, yield, and sensitivity of PCR were significantly improved by incorporating use of a thermostable DNA polymerase (2), resulting in a simplified procedure that has

Table 2. Quantitation of LCR amplification

Target molecules	Product formed, %*	Amplification†
6×10^7	134‡	
2×10^7	96	
6×10^6	107‡	
2×10^6	78	
6×10^5	85	
2×10^5	48	5.8×10^4
6×10^4	25	1.0×10^5
2×10^4	4.5	5.4×10^4
6×10^3	2.3	9.2×10^4
2×10^3	0.36	4.3×10^4
6×10^2	0.18	7.2×10^4
2×10^2	0.14	1.7×10^5
60 → 0§	<0.05	

Bands from 30-cycle LCR experiment described in Fig. 3 were excised from gels and assayed for radioactivity.

*Percentage product formed = cpm in product band/cpm in starting oligonucleotide band.

†Amplification = no. of product molecules formed/no. of target molecules.

‡At higher target concentration, DNA amplification was essentially complete after 20 cycles; slightly imprecise excision of 30-cycle bands from this portion of the gel probably accounts for product formed values $>100\%$.

§Product formed from 0 to 60 target molecules was indistinguishable from background (see Fig. 3).

Table 3. Quantitation of LCR amplification with or without mismatched competitor oligonucleotide

Target molecules	Complementary oligonucleotides	Complementary and mismatched oligonucleotides	
	Product formed, %*	Product formed, %*	Mismatched/complementary, %†
6×10^7 (β^A)	114*	93	1.0
2×10^7 (β^A)	93	95	1.8
6×10^6 (β^A)	102*	93	0.5
2×10^6 (β^A)	90	67	0.5
6×10^5 (β^A)	51	46	
2×10^5 (β^A)	31	23	
6×10^4 (β^A)	17	9.3	
2×10^4 (β^A)	8.6	2.9	
6×10^3 (β^A)	3.2	0.8	
0	<0.1	<0.1	
6×10^7 (β^S)	2.1	1.5	

One set of experiments contained 40 fmol each of β^A allele oligonucleotides 101 and 104 per tube, exactly as described for Fig. 3, whereas the second set had, in addition, 40 fmol each of oligonucleotides 103 and 106 per tube (forming G-T and C-A mismatches, respectively). Bands from 30-cycle LCR experiment, as described for Fig. 3, were excised from the gels and assayed for radioactivity.

*Percentage product formed = cpm in complementary product band/cpm in starting oligonucleotide band. Imprecise excision of two bands from the gel probably accounts for product formed values $>100\%$ (see Table 2).

†Percentage mismatched/complementary = cpm in bands of mismatched oligonucleotide products/cpm in band of complementary oligonucleotide products in same lane and indicates noise-to-signal ratio.

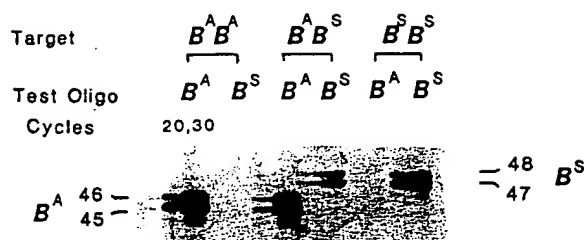


FIG. 4. Detection of β -globin alleles in human genomic DNA by autoradiogram. DNA was isolated from blood samples of normal ($\beta^A \beta^A$), carrier ($\beta^A \beta^S$), and sickle cell ($\beta^S \beta^S$) individuals as described. Genomic DNA (corresponding to 10 μ l of blood or $\approx 6 \times 10^4$ nucleated cells) was tested in two separate tubes containing labeled oligonucleotides (107 and 109; 200,000 cpm = 40 fmol each) and either unlabeled β^A test oligonucleotides (101 and 104) or unlabeled β^S test oligonucleotides (102 and 105; 40 fmol each). Both reaction mixtures were incubated under the same buffer (without salmon sperm DNA), enzyme, and cycle conditions described. Samples were electrophoresed, and the gel was autoradiographed overnight as described. Ligation products of 45 and 46 or 47 and 48 nucleotides indicate presence of the β^A - or β^S -globin gene, respectively. Oligo, oligonucleotide.

become widely applicable (23, 24). Similarly, this report demonstrates the utility of thermostable ligase for allele-specific gene detection under both LDR and LCR conditions. Both LCR and PCR amplification derive their specificity from the initial hybridization of primer to target DNA, and this is enhanced by (i) use of oligonucleotides of sufficient length to be unique in the human genome and (ii) use of temperatures near the oligonucleotide T_m . LCR amplification faithfully detected as few as 200 initial target molecules, as well as both β^A and β^S alleles directly from genomic DNA. LCR did not amplify a T-T, G-T, C-T, or C-A 3'-terminal mismatch, as has been reported for allele-specific PCR amplifications (25). Whether LCR will tolerate internal mismatches present in viral variants remains to be determined (25).

LCR amplification/detection is compatible with a primary amplification of genomic DNA by either PCR (2) or self-sustained sequence replication (3). Such a primary amplification could allow for LCR detection of emerging viral subpopulations where the mutations are known, such as the multiple mutations in human immunodeficiency virus conferring resistance to 3'-azido-3'-deoxythymidine (AZT) (26). One can also envisage multiplexing the primary amplification of dozens of loci simultaneously (27) and aliquoting products into separate microtiter wells. A subsequent round of LCR amplification/detection could then distinguish a particular target loci, even if it were initially amplified only in the amol range. Such a multiplex PCR/LCR detection assay, with the potential for an automated format, could (i) rapidly screen large populations for monogenic disease polymorphisms, (ii) distinguish several polymorphisms simultaneously from a single sperm to map the relative positions of these polymorphisms (28), and (iii) help eliminate current ambiguities in DNA identification of individuals for forensic or paternity cases (29).

The potential uses of thermostable enzymes that survive the temperature-cycling conditions required to denature double-stranded DNA are just now being tapped. With variations of the LCR concepts outlined above, thermostable ligase could be used to (i) covalently capture specific DNA fragments to a solid matrix, with the aid of "template oligonucleotides" (40- to 50-mers) complementary to both the fragment end as well as a second oligonucleotide attached to a solid support, (ii) covalently link PCR-generated fragments (for example, protein domains or exons) in specific order, and (iii) covalently link two members of a hexamer oligonucleotide library to form specific dodecamers for directed sequencing of cosmids and other large DNAs (30).

I thank Leroy Hood for valuable discussions and for suggesting cloning of thermostable ligase and its use in oligonucleotide ligation assays; Hamilton O. Smith, Tom Gingeras, and Marc Kahn for critical reading and discussions; and Kenneth Berns, Jef Boeke, Eric Spitzer, and John Zebala for discussions. I thank Deborah Nickerson for plasmids and discussions, Robert Kaiser and Suzanna Horvath for oligonucleotide synthesis, and Katherine Hajjar and Barbara Wall for providing blood samples from sickle cell patients. I thank Antje Koller and Jung Choi for expert technical assistance. This work was supported by grants from the National Institutes of Health (GM-41337-02) and the National Science Foundation (DMB-8714352).

- Landegren, U., Kaiser, R., Caskey, C. T. & Hood, L. (1988) *Science* 242, 229-237.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* 239, 487-491.
- Guatelli, J. C., Whitfield, K. M., Kwok, D. Y., Barringer, K. J., Richman, D. D. & Gingeras, T. R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874-1878.
- Wu, D. Y. & Wallace, R. B. (1989) *Genomics* 4, 560-569.
- Barringer, K., Orgel, L., Wahl, G. & Gingeras, T. R. (1990) *Gene* 89, 117-122.
- Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R. L., & Wallace, R. B. (1983) *Proc. Natl. Acad. Sci. USA* 80, 278-282.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. & Erlich, H. A. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6230-6234.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. (1989) *Proc. Natl. Acad. Sci. USA* 86, 2766-2770.
- Meyers, R. M., Larin, Z. & Maniatis, T. (1985) *Science* 230, 1242-1246.
- Cotton, R. G. H., Rodrigues, N. R. & Campbell, R. D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4397-4401.
- Kornher, J. S. & Livak, K. J. (1989) *Nucleic Acids Res.* 17, 7779-7784.
- Chehab, F. F. & Kan, Y. W. (1989) *Proc. Natl. Acad. Sci. USA* 86, 9178-9182.
- Landegren, U., Kaiser, R., Sanders, J. & Hood, L. (1988) *Science* 241, 1077-1080.
- Wu, D. Y. & Wallace, R. B. (1989) *Gene* 76, 245-254.
- Miyada, C. G. & Wallace, R. B. (1987) *Methods Enzymol.* 154, 94-107.
- Wilson, G. G. & Murray, N. E. (1979) *J. Mol. Biol.* 132, 471-491.
- Takahashi, M., Yamaguchi, E. & Uchida, T. (1984) *J. Biol. Chem.* 259, 10041-10047.
- Ishino, Y., Shinagawa, H., Makino, K., Tunasawa, S., Sakiyama, F. & Nakata, A. (1986) *Mol. Gen. Genet.* 204, 1-7.
- Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1074-1078.
- Higuchi, R. (1989) in *PCR Technology: Principles and Applications for DNA Amplification*, ed. Erlich, H. A. (Stockton, New York), p. 36.
- Horvath, S. J., Firca, J. R., Hunkapiller, T., Hunkapiller, M. W. & Hood, L. (1987) *Methods Enzymol.* 154, 314-326.
- Winslow, R. M. & Anderson, W. F. (1983) in *The Metabolic Basis of Inheritance*, eds. Stanbury, J. B., Wyngard, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 1666-1710.
- Erlich, H. A., ed. (1989) *PCR Technology: Principles and Applications for DNA Amplification* (Stockton, New York).
- Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic, New York).
- Kwok, S., Kellogg, D. E., Spasic, D., Goda, L., Levenson, C., Sninsky, J. J. (1990) *Nucleic Acids Res.* 18, 999-1005.
- Larder, B. A. & Kemp, S. D. (1989) *Science* 246, 1155-1158.
- Chamberlin, J. S., Gibbs, R. A., Rainer, J. E., Nguyen, P. N. & Caskey, C. T. (1988) *Nucleic Acids Res.* 16, 11141-11156.
- Cui, X., Li, H., Goradia, T. M., Lange, K., Kazazian, H. H., Jr., Galas, D. & Arnheim, N. (1989) *Proc. Natl. Acad. Sci. USA* 86, 9389-9393.
- Lander, E. (1989) *Nature (London)* 339, 501-504.
- Studier, F. W. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6917-6921.

Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations

(steroid 21-hydroxylase/heteroduplex/polymorphism)

RICHARD G. H. COTTON*†, NANDA R. RODRIGUES‡, AND R. DUNCAN CAMPBELL†

*Medical Research Council Immunochimistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom; and †Olive Miller Protein Laboratory, Murdoch Institute, Royal Childrens Hospital, Flemington Road, Parkville, Victoria 3052, Australia

Communicated by César Milstein, February 3, 1988 (received for review November 5, 1987)

ABSTRACT The chemical reactivity of thymine (T), when mismatched with the bases cytosine, guanine, and thymine, and of cytosine (C), when mismatched with thymine, adenine, and cytosine, has been examined. Heteroduplex DNAs containing such mismatched base pairs were first incubated with osmium tetroxide (for T and C mismatches) or hydroxylamine (for C mismatches) and then incubated with piperidine to cleave the DNA at the modified mismatched base. This cleavage was studied with an internally labeled strand containing the mismatched T or C, such that DNA cleavage and thus reactivity could be detected by gel electrophoresis. Cleavage at a total of

T and 21 C mismatches isolated (by at least three properly paired bases on both sides) single-base-pair mismatches was identified. All T or C mismatches studied were cleaved. By using end-labeled DNA probes containing T or C single-base-pair mismatches and conditions for limited cleavage, we were able to show that cleavage was at the base predicted by sequence analysis and that mismatches in a length of DNA could be readily detected by such an approach. This procedure may enable detection of all single-base-pair mismatches by use of sense and antisense probes and thus may be used to identify the mutated base and its position in a heteroduplex.

Definition of the exact single-base change in genes as a result of mutation is an important goal in genetic research. As sequencing complete genes to identify base changes is tedious, attempts have been made to improve the efficiency of the procedure (1-4). (i) Heteroduplexes formed between wild-type and variant DNAs have been treated with the single-strand-specific S1 nuclease to cleave the DNA at the point of the mismatched bases (1). (ii) The differential mobility of native and denatured DNA-DNA heteroduplexes coupled with their differential melting temperatures has been exploited by Myers *et al.* (2). (iii) Since this method was not generally applicable, Myers *et al.* (3) described a method in which mismatches in RNA-DNA heteroduplexes were cleaved by RNase A. (iv) An alternative approach in which RNase A was used to cleave mismatches in RNA-RNA heteroduplexes has also been described (4). (v) Novack *et al.* (5) have reported that single-base-pair mismatches in DNA-DNA heteroduplexes react with a carbodiimide.

As these methods did not detect all mutations, we have examined the chemical reactivity of mismatched bases in DNA-DNA heteroduplexes in more detail. We chose the steroid 21-hydroxylase (21-OHase) gene because of its medical importance and because of the large amount of polymorphism in the gene and pseudogene (6). We have screened those reagents, used first in the study of the secondary structure of tRNA (7) and then in DNA sequencing (8), that

lead to cleavage of the DNA chain on subsequent reaction with piperidine. Two reagents, osmium tetroxide and hydroxylamine, were found that potentially can recognize all variants, as they react with mismatched thymine (T) and cytosine (C), respectively.

MATERIALS AND METHODS

Preparation of DNA and Probes. Plasmid and M13 subclone DNA were prepared by standard methods (6). Internally labeled DNA probes were prepared from M13mp8 or M13mp9 subclones containing the DNA fragments that were used to generate the sequence of the 21-OHase A gene, the 21-OHase B gene and the mutant 21-OHase B gene (6) (Fig. 1).

Subclones carrying the desired DNA fragment complementary to the probe required were labeled by standard methods (9) by using the M13 universal sequencing primer. All dNTPs were at a concentration of 0.25 mM except dATP, which was added so that the [α - 32 P]dATP was diluted 1:9. Typically, 2 ng of primer was annealed with 50 ng of M13 DNA in 6.5 μ l, by heating at 90°C for 4 min followed by incubation at room temperature for 30 min. dNTPs were then added with 1 μ l of [α - 32 P]dATP (3000 Ci/mmol; 1 Ci = 37 GBq; Radiochemical Centre) and 1 μ l (7.5 units) of the Klenow fragment of DNA polymerase I (Pharmacia) in a final volume of 17.6 μ l, and the mixture was incubated at 20-24°C for 1 hr. All dNTPs were then added at 0.25 mM and incubated 30 min to chase. Samples were then extracted with chloroform/phenol, 1:1 (vol/vol), and the DNA was precipitated with ethanol. Immediately after labeling DNA was digested (in a final volume of 20-50 μ l) with restriction enzymes appropriate for the heteroduplex being studied (see Table 1). End-labeled DNA probes (see Table 1) were derived from the appropriate digests of the 3.7-kilobase *Taq* I fragment of the wild-type or mutant 21-OHase B genes or the 5.5-kilobase *Bgl* II-*Bam* HI fragment of the 21-OHase A gene, cloned in the *Pvu* II site of the plasmid pAT153/*Pvu* II/8 (6). Fragments were purified by electrophoresis in nondenaturing 4% polyacrylamide gels.

Heteroduplex Formation. Heteroduplexes contained unlabeled DNA from plasmid subclones digested with restriction enzymes appropriate for the heteroduplex being studied (see Table 1). A 12-fold molar excess of unlabeled digested plasmid DNA to labeled probe DNA was used for end-labeled probes, and a 50- or 200-fold molar excess of unlabeled digested plasmid DNA to labeled probe DNA was used for internally labeled probes. The mixture (20-100 μ l) was heated 5 min at 100°C and annealed 1 hr at 42°C in 0.3 M NaCl/3.5 mM MgCl₂/3 mM Tris-HCl, pH 7.7. Heteroduplex DNA was precipitated with ethanol and then taken up in

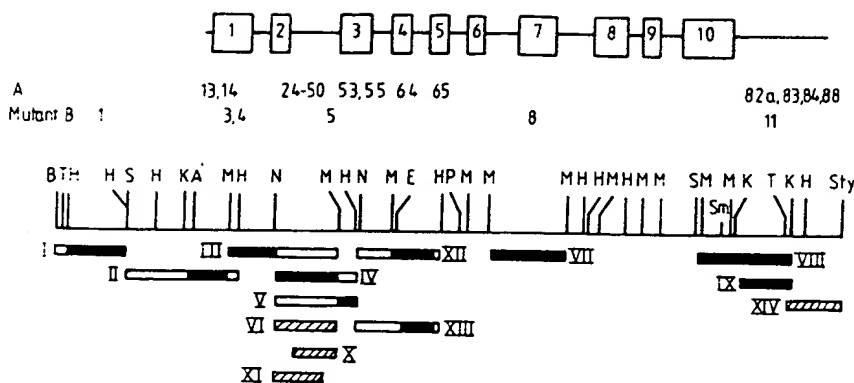


FIG. 1. Map of the wild-type 21-OHase B gene (6). The upper line shows the exon-intron structure of the gene, and the lower line is a restriction map. A, *Acc* I; B, *Bam*HI; E, *Eco*RI; H, *Hin*fI; K, *Kpn* I; M, *Msp* I; N, *Nco* I; P, *Pvu* II; S, *Sst* I; Sm, *Sma* I; Sty, *Srv* I; T, *Taq* I. The numbers above the restriction map show the approximate positions of differences of the mutant B gene (6) (lower row of numbers) and the A gene (upper row of numbers) from the wild-type B gene with the numbering starting from the first difference from the 5' end. Only some of those differences studied are shown. The horizontal bars represent the B gene, the mutant B gene, or the A gene DNA from M13 subclones used as specified in Table 1 or from pAT plasmid DNA used for end-labeling studies (Table 1). The solid boxes represent the DNA used as probes. The hatched boxes represent the pAT plasmid DNA used for the end-labeling studies. The Roman numerals represent the regions used to produce probes and are referred to in Table 1.

distilled water at 1000 cpm/ μ l. Approximately 6000 cpm of labeled probe DNA was used per assay tube.

Hydroxylamine Treatment of DNA. Hydroxylamine hydrochloride (1.39 g) (Analar grade; BDH) was dissolved in 1.6 ml of distilled water and the pH was adjusted to 6.0 with diethylamine (Fluka). The final volume was \approx 4 ml, giving a hydroxylamine concentration of \approx 2.5 M.

DNA in 6 μ l of distilled water was treated with 20 μ l of hydroxylamine solution at 37°C for 2 hr or as indicated. The reaction was stopped by transferring the mixture to ice and adding 200 μ l of stop solution containing 0.3 M sodium acetate, 0.1 mM Na_2EDTA (pH 5.2) and tRNA (25 μ g/ml; baker's yeast, Boehringer Mannheim), and the DNA was precipitated with ethanol. After a further ethanol precipitation, the DNA pellet was washed once with 70% (vol/vol) ethanol and dried.

Osmium Tetroxide Treatment of DNA. DNA in 6 μ l of distilled water was treated with 15 μ l of 4% (wt/vol) osmium tetroxide in water (Aldrich) in a total volume of 24.5 μ l containing 1 mM EDTA, 10 mM Tris-HCl (pH 7.7), and 1.5% (vol/vol) pyridine. Incubation was at 37°C as indicated. The reaction was stopped as described for hydroxylamine.

Piperidine Cleavage. Chemical cleavage of the C and T bases that had reacted with hydroxylamine or osmium tetroxide was achieved by incubating the heteroduplexes with piperidine (8). Piperidine (50 μ l at 1 M) was added to the dry DNA pellet and incubated at 90°C for 30 min. DNA was precipitated with ethanol, washed with 70% (vol/vol) ethanol, and dried. For osmium tetroxide-treated DNA, ethanol precipitation after piperidine treatment was in a dry ice/methanol bath, and all subsequent operations were at or below 4°C until the dried pellet was obtained.

Electrophoresis of Products. Samples were incubated in 10 μ l of 60% (vol/vol) formamide/0.1% bromophenol blue/0.1% xylene cyanol FF/35 mM Na_2EDTA , pH 7.4, at 100°C for 4 min before application to denaturing urea gels (8). Cleavage and recovery was estimated by measuring radioactivity in gel slices and is reported at 2 hr for hydroxylamine cleavage and at 30 min for osmium tetroxide. Recovery was calculated relative to an unincubated control.

RESULTS

Hydroxylamine Cleavage of Mismatched C Bases. Preliminary experiments indicated that optimal cleavage of mismatched C bases was obtained with a 2-hr incubation in 2 M hydroxylamine at pH 6. Lower concentrations were not as effective, and longer times led to the destruction of the DNA.

The cleavage of C-C, C-T, and C-A mismatches [mutations B8, B4, and B11, respectively (6)] as a function of time is shown in Fig. 2. Cleavage at 2 hr was 93%, 88%, and 74% with recoveries of 65%, 71%, and 23%, respectively (recoveries of 70% were later consistently achieved by use of methanol/dry ice for ethanol precipitation). The 215-base fragment in Fig. 2B is due to cleavage of a C-A mismatch (mutation B3) that lies 20 bases from the C-T mismatch. Cleavage at this C-A mismatch was not quantitated. In all cases the size of the cleavage products was consistent with cleavage at the respective mismatches. Controls—heteroduplexes with no incubation, with no hydroxylamine added, or with no piperidine added, a homoduplex with the same labeled strand, and a heteroduplex with the opposite strand labeled—showed no specific cleavage of the labeled strand (Fig. 2). For mutation B11 (C-A mismatch), the probe included 10 bases of the vector and the size heterogeneity seen in Fig. 2C (lane 5) is due to cleavage of those unpaired bases.

The above results are consistent with cleavage at the position of the mismatch identified by sequence analysis (6) but do not prove that it is at this point. To determine the exact position of cleavage the 3' end of probe VI was end-labeled by using the Klenow fragment of DNA polymerase I. A portion of the probe preparation was sequenced by the Maxam-Gilbert method (8). Another portion of the preparation was used to make heteroduplex that was then incubated with hydroxylamine for 2 hr. The product of the cleavage reaction is exactly adjacent to the mismatched C (Fig. 3, lane 5). Two minor products, 1 and 3 bases from the mismatched C, are also apparent presumably due to propagation (10) where paired C bases near the mismatch show some reactivity with hydroxylamine, similar to unmatched bases near the loops of tRNA.

The ability of hydroxylamine and piperidine to cleave C mismatches in various contexts is summarized in Table 1. Cleavages of 90%, 84%, and 87% were observed for a C-C mismatch (mutation B5) and two C-A mismatches (mutations B1 and A82a), respectively. In other cases only the ability to cleave was recorded.

Screening of a larger number of mismatches for cleavage was possible in the probe IV/V region (see Fig. 1) making use of the large number of differences between the 21-OHase A and B genes in this area. An end-labeled probe was used to facilitate the positioning of cleavage and partial cleavage with hydroxylamine was used to increase the yield of the various bands expected. One such experiment is illustrated in Fig. 3 (lane 6) with a probe from region VI that hybridized to the

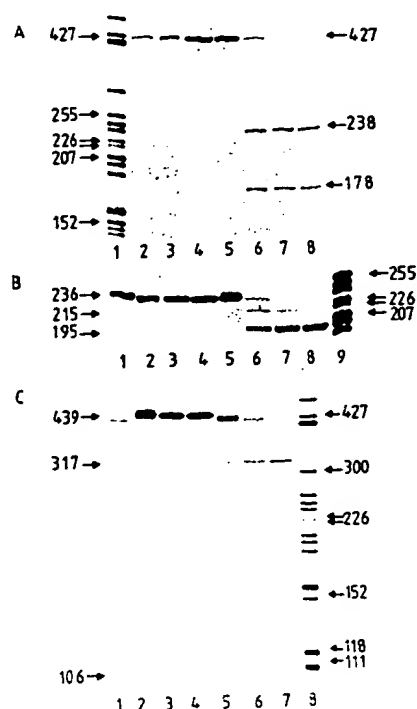


Fig. 2. Hydroxylamine reaction with an internally labeled probe. (A) C-C mismatch (mutation B8). (B) C-A and C-T mismatches (mutations B4 and B3, respectively). (C) C-A mismatch (mutation B11). All incubations were at 37°C with 2 M hydroxylamine unless otherwise indicated. (A) Heteroduplexes (1.2 μ g) with C-C as the only mismatch were incubated for 0 (lane 5), 30 (lane 6), 60 (lane 7), or 120 (lane 8) min. Controls were homoduplexes (1.2 μ g) with the same labeled strand but with an unlabeled wild-type DNA incubated 120 min (lane 2), heteroduplexes (1.2 μ g) incubated for 120 min without the subsequent addition of piperidine (lane 3), and heteroduplexes (1.2 μ g) incubated without hydroxylamine for 120 min (lane 4). Lane 1 contains molecular size markers. Size markers in bases are on the left; sizes of fragments and original probes are on the right. (B) Heteroduplexes (2 μ g) containing C-A and C-T as the only mismatches were incubated for 0 (lane 3), 30 (lane 6), 60 (lane 7), and 120 (lane 8) min. The band at 195 bases represents cleavage at the C-T mismatch, and the band at 215 bases represents cleavage at the C-A mismatch only. Controls were the complementary heteroduplexes (2 μ g) with a G-T mismatch (in which the 5' mutant strand is labeled and the wild-type DNA is unlabeled) incubated for 0 (lane 1) and 120 (lane 2) min, heteroduplexes (2.0 μ g) incubated for 120 min without hydroxylamine (lane 4), and heteroduplexes (2.0 μ g) incubated for 120 min without subsequent addition of piperidine (lane 5). Lane 9 contains molecular size markers. (C) Heteroduplexes (0.3 μ g) containing C-A as the only mismatch (except for ragged ends from a small region of the M13 vector, see Table 1) were incubated for 0 (lane 4), 30 (lane 5), 60 (lane 6), or 120 (lane 7) min. Controls were homoduplexes (0.3 μ g) with the same labeled strand but with an unlabeled wild-type DNA (Table 1) incubated for 120 min (lane 1), heteroduplexes (0.3 μ g) incubated for 120 min without subsequent addition of piperidine (lane 2), and heteroduplexes (0.3 μ g) incubated for 120 min without hydroxylamine (lane 3). Lane 8 contains molecular size markers. For B and C, numbers on the right refer to size markers in bases and numbers on the left are the size of fragments and original probes.

unlabeled 21-OHase A gene DNA. Labeling was at the 3' end of the sense strand. After partial (30 min) reaction with hydroxylamine only the expected 246-, 147-, 90-, 62-, and 36-base fragments are seen. This is consistent with cleavage at the five expected C mismatches: C-A, C-T, C-T, C-C, and C-C (mutations A24, A35, A45, A48, and A50, respectively). One mutation (A48) did not produce an isolated C mismatch due to the presence of a T-C mismatch immediately adjacent.

Use of a probe from region X (Fig. 1) potentially allows the study of 8 C mismatches and 1 unpaired C base, whereas use

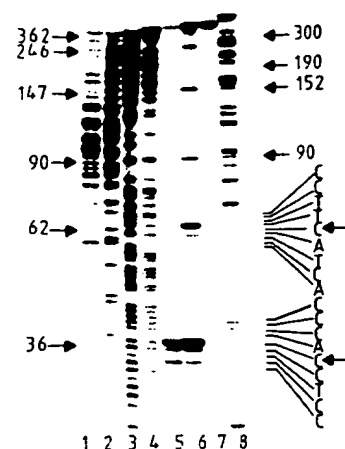


Fig. 3. Analysis of the position of cleavage of a heteroduplex by hydroxylamine with end-labeled DNA. Lanes 1-4 show a Maxam-Gilbert sequencing ladder of the end-labeled probe (region VI) (G, A, T, and C, respectively). Lane 5 shows the same labeled DNA after formation of a heteroduplex with a 12 times excess of unlabeled mutant DNA (Table 1) and incubation with 2 M hydroxylamine for 2 hr at 37°C. Lane 6 shows the same labeled DNA after formation of a heteroduplex with a 12 times excess of unlabeled A gene DNA and incubation with 2 M hydroxylamine for 30 min at 37°C. Lane 7 shows the same labeled DNA after formation of a homoduplex with a 12 times excess of unlabeled wild-type DNA and incubation with 2 M hydroxylamine for 2 hr at 37°C. To the right are shown the sequences around two of the cleavage points (arrows) where the sequence is readable. Numbers on the left are the sizes of fragments (and probe) produced by the heteroduplex shown in lane 6. Numbers on the right are the sizes of markers (contained in lane 8). The amount of DNA in lanes 5-7 was 2.2 μ g.

of a probe from region XI (Fig. 1) potentially allows the study of 11 C mismatches and 1 unpaired C base in a loop. Table 1 shows those cleavages where neighboring mismatches are >3 bases away. With the region X and the region XI probes in the regions able to be assessed, all C mismatches or unpaired C bases were cleaved. Except for those C bases near the C mismatches that showed lesser cleavage, presumably due to propagation, no unexpected cleavages were found.

Osmium Tetroxide Cleavage of Mismatched T Bases. The cleavage of T-G, T-C, and T-T mismatches (mutations B3, B4, and A64, respectively) are shown with increasing time in Fig. 4. Cleavage was 61%, 78%, and 17% with recoveries of 33%, 30%, and 21%, respectively. Heteroduplex controls without incubation, osmium tetroxide, or piperidine, or a homoduplex control showed no specific cleavage. In all cases the size of the cleavage products was consistent with cleavage at the respective mismatch site.

Table 1 summarizes the results obtained with a further three T mismatches studied with an internally labeled probe. Substantial cleavage of the probes used was observed for T-G (mutation B10a) and for T-C (mutation A65) mismatches. Cleavage of T mismatches that were not quantitated were studied with end-labeled probes (see below). Cleavage at T-G and T-C mismatches (mutations A17, A23, A29, and A30) is shown in Fig. 5 and Table 1. T-G mismatches (mutations A46, A84, and A88) and a T-T mismatch (mutation A31) were cleaved in other end-labeled probes (Table 1).

To determine the exact position of cleavage, products of the Maxam-Gilbert sequencing reactions of the end-labeled probe was electrophoresed next to the heteroduplex that had reacted with osmium tetroxide (Fig. 5, lanes 1-6). It can be seen that the two isolated T mismatches (A29 and A30) are cleaved at the position of the mismatch.

Fig. 5 also shows the use of end-labeled probe in a heteroduplex with unlabeled DNA suspected of containing sequence changes; partial cleavage is a convenient method

Table 1. Summary of C and T mismatches cleaved

Mismatch	Mutation	Probe*	Cleavage†	Sequences at mismatch‡
C/A	B3	5'/III/M	+ve(+ve)	TGCAC(C)TGCTG
C/T	B4		88(57)	CTCCC(C)ATCTA
C/C	B8	3'/VII/B	93(79)	CCATG(C)TCGGC
C/A	B11	5'/VIII/B	74(+ve)	CTCGG(C)AGTCA
C/C	B5	5'/IV/B	90	CCCCA(C)CTCCT
C/A	A13		+ve	CGTCT(C)GCCAT
C/A	A14	5'/II/B	+ve	CCTCC(C)GCCTC
C/A	B1	3'/I/M	84(81)	CCTCC(C)TTTAC
C/A	A82a	5'/IX/A	87	GCTCC(C)GTACG
C/A	A24	5'/VI/B	E	CTATG(C)TGCCC
C/T	A35	5'/VI/B	E	AGGTC(C)CTGGA
C/T	A45	5'/VI/B	E	GAGGC(C)GAAGA
C/C	A48	5'/VI/B	E	GCCTT(C)ATCAG
C/C	A50	5'/VI/B	E	CCCCA(C)CTCCT
C/T	A43	3'/X/B	E	CCAAC(C)CCTGC
C/A	A44	3'/X/B	E	CCTCC(C)CAACC
C/T	A47	3'/X/B	E	GAAGG(C)AGCTG
C/T	A23	3'/XI/A	E	CAAGA(C)CCCAT
C/A	A25	3'/XI/A	E	GAATT(C)AAGAC
C/A	A26	3'/XI/A	E	GAGAC(C)AGGAA
C/A	A27	3'/XI/A	E	GATCA(C)TTGAG
C/A	A28	3'/XI/A	E	GAGGC(C)GAGGT
C/T	A29	3'/XI/A	E	GGCTC(C)CACTT
T/T	A64	5'/XII/B	17	CATCA(T)CTGTT
T/C	B4	3'/III/B	78	TAGAT(T)GGGAG
T/G	B3	5'/III/B	61	TGCAC(T)TGCTG
T/G	B10a		46	GCTCC(T)GTACG
C/A	B11	5'/VIII/B	(+ve)	CTCGG(C)AGTCA
T/C	A65		57	AAGGA(T)GGAGT
C/C	A67	3'/XIII/A	(+ve)	TTGAC(C)TCCTG
T/G	A30	5'/VI/B	E	ACCTT(T)GGGGC
T/C	A29	5'/VI/B	E	AAGTG(T)GAGCC
T/C	A23	5'/VI/B	E	ATGGG(T)TCTTG
T/G	A17	5'/VI/B	E	AGGGC(T)GGGGG
T/T	A31	3'/X/B	E	GGGGA(T)GCCCC
T/G	A46	3'/X/B	E	GCAGC(T)GAGGG
T/G	A88	5'/XIV/B	E	TTAAT(T)CTGAG
T/G	A84	5'/XIV/B	E	TGCTC(T)TCCCG
C/A	A89	5'/XIV/B	(E)	GCTGG(C)CCTTT

*Entry gives the probe sense, area of the gene used, and the gene used (see Fig. 1).

†Cleavage is defined in *Materials and Methods*. +ve indicates cleavage seen but not quantitated. E denotes that end-labeled probe was used and cleavage was seen but not quantitated. Value in parentheses refers to cleavage of C mismatches with osmium tetroxide.

‡Mismatched bases that are cleaved are in parentheses and nearby mismatched bases are underlined.

*Mutations were described after publication of ref. 6.

for detecting these differences. After a short incubation with osmium tetroxide (lane 5, 1 min; lane 6, 5 min) and subsequent cleavage of the heteroduplex with piperidine, a number of bands not seen in the homoduplex control treated in the same way (lane 8) are apparent. Consideration of the sequencing tracks (lanes 1–4), the molecular size markers (lane 9), and the sequence allows assignment of the bands to specific T bases. The five single-base-pair T mismatches are indicated by mutation name and lead to five of the six major bands seen in lane 5. The sixth major band (second from top) results from the cleavage of a T next to a loop in the 21-OHase A gene due to a 4-base insertion. The next strongest bands, two below mutation A30, are due to cleavage of T mismatches next to a single-base-pair mismatch (mutation A31) or 3 bases from a 3-base insert. Three further examples of the former are seen in the three faint bands above mutation A29.

The minor band below mutation A23 (lanes 5 and 6) and at the second hydroxylamine cleavage of C from the bottom (lane 7) is consistent with slower rate of reaction of osmium tetroxide with C mismatches relative to T mismatches (see below).

The hydroxylamine cleavage of the same heteroduplex (lane 7) (also illustrated in Fig. 3, lane 6) illustrates how a stretch of DNA can be scanned for all T and C mismatches.



FIG. 4. Osmium tetroxide reaction with internally labeled probe. (A) T-G mismatch (mutation B3). (B) T-C mismatch (mutation B4). (C) T-T mismatch (mutation A64). All incubations were at 37°C with 2.4% (wt/vol) osmium tetroxide at 37°C unless otherwise indicated. (A) Heteroduplexes (0.08 μ g) containing T-G and A-G as the only mismatches were incubated for 0 (lane 4), 30 (lane 5), 60 (lane 6), or 120 (lane 7) min. Controls were homoduplexes (0.08 μ g) with the same labeled strand but with unlabeled wild-type DNA incubated for 120 min with (lane 2) or heteroduplexes without (lane 3) osmium tetroxide. (B) Heteroduplexes (0.35 μ g) containing T-C and A-C as the only mismatches were incubated for 0 (lane 5), 15 (lane 6), 30 (lane 7), or 60 (lane 8) min. Controls were homoduplexes (0.35 μ g) with the same labeled strand but with unlabeled wild-type DNA incubated for 30 min (lane 2), heteroduplexes (0.35 μ g) incubated for 50 min without the subsequent addition of piperidine (lane 3), and heteroduplexes (0.35 μ g) incubated at 37°C for 60 min without osmium tetroxide (lane 4). (C) Heteroduplexes (0.19 μ g) containing T-T as the only mismatch were incubated for 0 (lane 5), 15 (lane 6), 30 (lane 7), or 60 (lane 8) min. Controls, incubated for 60 min, were homoduplexes (0.19 μ g) with the same labeled strand but unlabeled wild-type DNA (lane 2), heteroduplexes (0.19 μ g) incubated without the subsequent addition of piperidine (lane 3), and heteroduplexes (0.19 μ g) incubated without osmium tetroxide (lane 4). In B and C only half the DNA was loaded onto lanes 3 and 4. For A–C, numbers on the left refer to the size of the marker fragments (lanes 1) in bases and numbers on the right refer to the size of the fragments and original probe in bases.

Osmium Tetroxide Cleavage of Mismatched C Bases. The C mismatches studied for cleavage by hydroxylamine and piperidine were also studied for cleavage by osmium tetroxide and piperidine (Table 1). By using internally labeled probes containing C-T, C-C, and C-A mismatches (mutations B4, B8, and B11), cleavages of 57%, 78%, and 81%, respectively, were found. A further two C-A mismatches (mutations B3 and B11) were also cleaved, but the values were not quantifiable. The rate of cleavage of the C mismatches by osmium tetroxide was slower than cleavage of T mismatches (data not shown).

DISCUSSION

We have screened a variety of reagents for their ability to react with purine or pyrimidine bases when they are mismatched in a duplex in such a way that the probe containing the mismatched bases is cleaved at that point by piperidine. Such reagents included hydrazine, potassium permanganate, formic acid, sodium hydroxide, diethyl pyrocarbonate, methylene blue, hydroxylamine, and osmium tetroxide and have



FIG. 5. Analysis of the position of cleavage of the heteroduplex shown in Fig. 3 by osmium tetroxide as well as hydroxylamine with end-labeled DNA. Lanes 1-4 show a Maxam-Gilbert sequencing ladder of the end-labeled probe (region VI) (G, A, T, and C, respectively). Lanes 5 and 6 show the same labeled DNA after formation of a heteroduplex with a 12 times excess of unlabeled A gene DNA (Table 1) and incubation with 2.4% (wt/vol) osmium tetroxide for 1 and 5 min, respectively, at 37°C. Lane 7 shows the same DNA heteroduplex treated with 2 M hydroxylamine for 10 min at 37°C. Lane 8 is a homoduplex control with the same labeled strand but with wild-type unlabeled DNA treated with 2.4% (wt/vol) osmium tetroxide for 5 min at 37°C. To the left are shown the sequences around two of the cleavage points (arrow), where the sequence is readable. Letters and numbers on the left are the mutation numbers represented by the cleavages in lanes 5 and 6, and numbers on the right are the size in bases of the markers (lane 9). The amount of DNA in lanes 5-8 was 2.6 µg.

been used for structural studies of tRNA (7, 10), for sequencing nucleic acids (8), and for Z-DNA studies (11). In initial experiments, hydroxylamine (12-14) and osmium tetroxide (12), showed considerable promise, and conditions were established for maximal cleavage of mismatched C and T, respectively (data not shown). We applied these conditions to a large number of T and C mismatches and showed that all 13 T mismatches studied were cleaved, including 2 T-T, 4 T-C, and 7 T-G mismatches. All 21 C mismatches studied were also cleaved, including 2 C-C, 7 C-T, and 12 C-A mismatches. At least one example of each C mismatch was cleaved with osmium tetroxide at a slower rate consistent with earlier studies (15). Previous work on tRNA with osmium tetroxide (16) and *O*-methylhydroxylamine (17), a compound related to hydroxylamine, allowed us to predict that unmatched C or T would be reactive, and this was found in three cases of unmatched C bases (data not shown). Thus all types of mutations (i.e., insertions, deletions, and base changes) can be detected by the procedure described here.

The use of end-labeled probes (Figs. 3 and 5) allowed us (i) to confirm that, for selected cases, the point of cleavage was at the point predicted; (ii) to collect further examples of C or T mismatch cleavages; and (iii) to use the above findings to test procedures to detect mismatches, and hence mutations or polymorphisms, after wild-type (or reference) DNA had been annealed to variant DNA.

The method for detection of mismatched bases described in this paper as applied to cloned DNA can be contrasted with two other methods, the RNase method (3, 4) and the carbodiimide method (5). The RNase method (3, 4) needs an extra step of cloning (into the SP6 vector) beyond that needed for the carbodiimide method (5) or the method described here. However, the greatest drawback of the RNase method appears to be the variable cleavage of some mismatches from 0 of 6 G-C mismatches to 1 of 14 G-T and 1 of 7 G-A mismatches to excellent cleavage of all 22 C-A mismatches (3). The study of mismatches with the end-labeled probe is theoretically possible with the RNase method but has not yet been reported. The carbodiimide method requires the heteroduplex first to be made blunt-ended, but its potential for detecting mismatches is unclear; the results for only two mismatches (T-C and G-T) were given, although positive results for G-G and T-T mismatches were mentioned but not shown. Because this method is not a cleavage method, fragments cannot be detected by PAGE. In contrast, the strengths of our method appear to be (i) that no extra cloning is required beyond that for cloning and sequencing the wild-type (reference) DNA, (ii) that it is a cleavage method that allows easy assessment, (iii) that because it is a chemical method, it may be more reproducible than enzymatic methods, (iv) that comparison with a Maxam-Gilbert sequencing ladder of limited cleavage of a heteroduplex allows rapid and ready identification of position and type of mismatch, and (v) that as no mismatches have yet been found that do not cleave, it is possible that all mismatches may be detectable. Thus if a labeled probe contains a mismatched T or C in its heteroduplex, it can be readily detected, although a mismatched A or G will not. However, a probe of the opposite sense will contain a mismatched T or C, respectively, and the mismatch then will be detected.

Our method should be applicable to genomic DNA, particularly in the analysis of the defective genes associated with inherited diseases and in the study of oncogenes that differ by a few bases. In addition our procedure may be used to compare related virus isolates and may provide a convenient and rapid check of *in vitro*-mutagenized DNA fragments.

We thank Ms. Carolyn Brooks and Mr. Frank Caddick for their help in preparation of the manuscript and Mrs. Sandra Smith for technical assistance. This work was supported by the Nuffield Foundation and the National Health and Medical Research Council of Australia. N.R.R. is a Rhodes Scholar.

- Shenk, T. E., Rhodes, C., Rigby, P. W. J. & Berg, P. (1975) *Proc. Natl. Acad. Sci. USA* 72, 989-993.
- Myers, R. M., Lumelsky, N., Lerman, L. S. & Maniatis, T. (1985) *Nature (London)* 313, 495-498.
- Myers, R. M., Larin, Z. & Maniatis, T. (1985) *Science* 230, 1242-1246.
- Winter, E., Yamamoto, F., Almoguera, C. & Perucho, M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7575-7579.
- Novack, D. F., Casna, N. J., Fischer, S. G. & Ford, J. P. (1986) *Proc. Natl. Acad. Sci. USA* 83, 586-590.
- Rodriguez, N. R., Dunham, I., Yu, C. Y., Carroll, M. C., Porter, R. R. & Campbell, R. D. (1987) *EMBO J.* 6, 1653-1661.
- Cramer, F. (1971) *Prog. Nucleic Acid Res. Mol. Biol.* 11, 391-421.
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- Furlong, J. C. & Lilley, D. M. J. (1986) *Nucleic Acids Res.* 14, 3995-4007.
- Johnston, B. H. & Rich, A. (1985) *Cell* 42, 713-724.
- Kochetkov, N. K. & Budovskii, E. I. (1972) *Organic Chemistry of Nucleic Acids* (Plenum, London), Part B.
- Morozova, T. M. & Salganik, R. I. (1964) *Biokhimiya* 29, 17.
- Rubin, C. M. & Schmid, C. W. (1980) *Nucleic Acids Res.* 8, 4613-4619.
- Burton, K. & Riley, W. T. (1966) *Biochem. J.* 98, 70-77.
- Lilley, D. M. J. & Palecek, E. (1984) *EMBO J.* 3, 1187-1192.
- Cashmore, A. R., Brown, D. M. & Smith, J. D. (1971) *J. Mol. Biol.* 59, 359-373.

Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics

William E. Evans* and Mary V. Relling

Genetic polymorphisms in drug-metabolizing enzymes, transporters, receptors, and other drug targets have been linked to interindividual differences in the efficacy and toxicity of many medications. Pharmacogenomic studies are rapidly elucidating the inherited nature of these differences in drug disposition and effects, thereby enhancing drug discovery and providing a stronger scientific basis for optimizing drug therapy on the basis of each patient's genetic constitution.

There is great heterogeneity in the way individuals respond to medications, in terms of both host toxicity and treatment efficacy. Potential causes for such variability in drug effects include the pathogenesis and severity of the disease being treated; drug interactions; and the individual's age, nutritional status, renal and liver function, and concomitant illnesses. Despite the potential importance of these clinical variables in determining drug effects, it is now recognized that inherited differences in the metabolism and disposition of drugs, and genetic polymorphisms in the targets of drug therapy (such as receptors), can have an even greater influence on the efficacy and toxicity of medications. Clinical observations of such inherited differences in drug effects were first documented in the 1950s, exemplified by the relation between prolonged muscle relaxation after suxamethonium and an inherited deficiency of plasma cholinesterase (1), hemolysis after antimalarial therapy and the inherited level of erythrocyte glucose 6-phosphate dehydrogenase activity (2), and peripheral neuropathy of isoniazid and inherited differences in acetylation of this medication (3). Such observations gave rise to the field of "pharmacogenetics," which focuses largely on genetic polymorphisms in drug-metabolizing enzymes and how this translates into inherited differences in drug effects [reviewed in (4)].

The molecular genetic basis for these inherited traits began to be elucidated in the late 1980s, with the initial cloning and characterization of a polymorphic human gene encoding the drug-metabolizing enzyme debrisoquin hydroxylase (*CYP2D6*) (5). Genes are considered functionally "polymorphic" when allelic variants exist stably in the population, one or more of which alters the activity of the encoded protein in relation to the wild-type

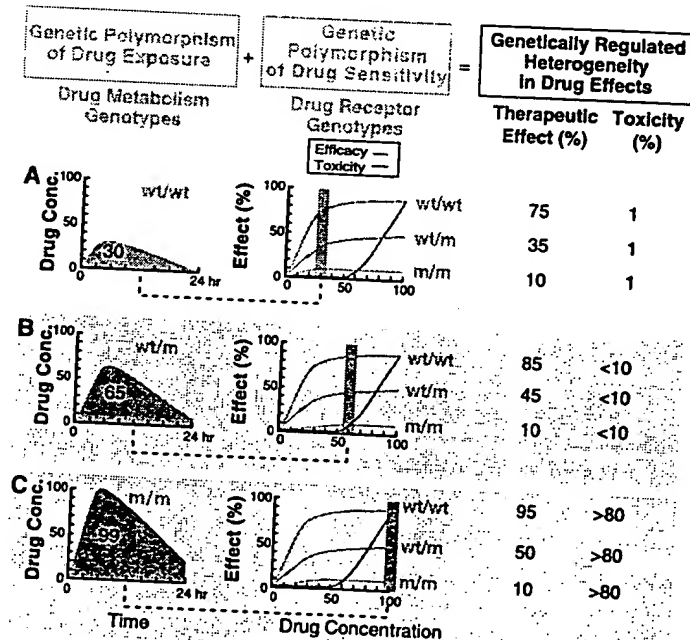
sequence. In many cases, the genetic polymorphism is associated with reduced activity of the encoded protein, but there are also examples where the allelic variant encodes proteins with enhanced activity. Since the cloning and characterization of *CYP2D6*, human genes involved in many such pharmacogenetic traits have been isolated, and their molecular mechanisms have been elucidated, and their clinical importance has been more clearly defined. Inherited differences in drug-metabolizing capacity are generally monogenic traits, and their influence on the pharmacoki-

netics and pharmacologic effects of medications is determined by their importance for the activation or inactivation of drug substrates. The effects can be profound toxicity for medications that have a narrow therapeutic index and are inactivated by a polymorphic enzyme (for example, mercaptopurine, azathioprine, thioguanine, and fluorouracil) (6) or reduced efficacy of medications that require activation by an enzyme exhibiting genetic polymorphism (such as codeine) (7).

However, the overall pharmacologic effects of medications are typically not monogenic traits; rather, they are determined by the interplay of several genes encoding proteins involved in multiple pathways of drug metabolism, disposition, and effects. The potential polygenic nature of drug response is illustrated in Fig. 1, which depicts the hypothetical effects of two polymorphic genes: one that determines the extent of drug inactivation and

Fig. 1. Polygenic determinants of drug effects. The potential consequences of administering the same dose of a medication to individuals with different drug-metabolism genotypes and different drug-receptor genotypes is illustrated. Active drug concentrations in systemic circulation are determined by the individual's drug-metabolism genotype (green lettering), with (A) homozygous wild type (wt/wt) patients converting 70% of a dose to the inactive metabolite, leaving 30% to exert an effect on the target receptor. (B) For the patient with heterozygous (wt/m) drug-metabolism genotype, 35% is inactivated,

whereas (C) the patient with homozygous mutant (m/m) drug metabolism inactivates only 1% of the dose by the polymorphic pathway, yielding the three drug concentration-time curves. Pharmacological effects are further influenced by different genotypes of the drug receptor (blue lettering), which have different sensitivity to the medication, as depicted by the curves of drug effect at any given drug concentration in comparison to those with a wt/wt receptor genotype, whereas those with m/m receptor genotypes are relatively refractory to drug effects at any plasma drug concentration. These two genetic polymorphisms (in drug metabolism and drug receptors) toxicity) ranges from a favorable 75 in the patient with wt/wt genotypes for drug metabolism and drug receptors to <0.13 in the patient with m/m genotypes for drug metabolism and drug



Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA, and Colleges of Pharmacy and Medicine, University of Tennessee, Memphis, TN 38105, USA.

*To whom correspondence should be addressed. E-mail: william.evans@stjude.org

another that determines the sensitivity of the drug receptor. The polymorphic drug-metabolizing enzyme, which exhibits codominant inheritance (that is, three phenotypes), determines the plasma concentrations to which each individual is exposed, whereas the polymorphic receptor determines the nature of response at any given drug concentration. This example assumes that drug toxicity (Fig. 1, red lines) is determined by nonspecific effects or through receptors that do not exhibit functionally important genetic polymorphisms, although clearly toxicity can also be determined by genetic polymorphisms in drug receptors. Thus, the individual with homozygous wild-type drug-metabolizing enzymes and drug receptors (Fig. 1A) would have a high probability of therapeutic efficacy and a low probability of toxicity, in contrast to an individual with homozygous mutant genotypes for the drug-metabolizing enzyme and the drug receptor, in which the likelihood of efficacy is low and that of toxicity is high (Fig. 1C).

Such polygenic traits are more difficult to elucidate in clinical studies, especially when a medication's metabolic fate and mechanisms of action are poorly defined. However, biomedical research is rapidly defining the molecular mechanisms of pharmacologic effects, genetic determinants of disease pathogenesis, and functionally important polymorphisms in genes that govern drug metabolism and disposition. Moreover, the Human Genome Project, coupled with functional genom-

ics and high-throughput screening methods, is providing powerful new tools for elucidating polygenic components of human health and disease. This has spawned the field of "pharmacogenomics", which aims to capitalize on these insights to discover new therapeutic targets and interventions and to elucidate the constellation of genes that determine the efficacy and toxicity of specific medications. In this context, pharmacogenomics refers to the entire spectrum of genes that determine drug behavior and sensitivity, whereas pharmacogenetics is often used to define the more narrow spectrum of inherited differences in drug metabolism and disposition, although this distinction is arbitrary and the two terms are now commonly used interchangeably. Ultimately, knowledge of the genetic basis for drug disposition and response should make it possible to select many medications and their dosages on the basis of each patient's inherited ability to metabolize, eliminate, and respond to specific drugs. Herein, we provide examples that illustrate the current status of such pharmacogenomic research and discuss the prospects for near-term advances in this field.

Genetic Polymorphisms in Drug Metabolism and Disposition

Until recently, clinically important genetic polymorphisms in drug metabolism and disposition were typically discovered on the basis of phenotypic differences among individuals in the population (8), but the framework for discovery of pharmacogenetic traits is

rapidly changing. With recent advances in molecular sequencing technology, gene polymorphisms [such as single-nucleotide polymorphisms (SNPs), and especially SNPs that occur in gene regulatory or coding regions (cSNPs)] may be the initiating discoveries, followed by biochemical and, ultimately, clinical studies to assess whether these genomic polymorphisms have phenotypic consequences in patients. This latter framework may permit the elucidation of polymorphisms in drug-metabolizing enzymes that have more subtle, yet clinically important consequences for interindividual variability in drug response. Such polymorphisms may or may not have clear clinical importance for affected medications, depending on the molecular basis of the polymorphism, the expression of other drug-metabolizing enzymes in the patient, the presence of concurrent medications or illnesses, and other polygenic clinical features that impact upon drug response. In Fig. 2, we have highlighted those drug-metabolizing enzymes known to exhibit genetic polymorphisms with incontrovertible clinical consequences; however, almost every gene involved in drug metabolism is subject to common genetic polymorphisms that may contribute to interindividual variability in drug response. Table 1 provides examples of how these genetic polymorphisms can translate into clinically relevant inherited differences in drug disposition and effects, a comprehensive summary of which is available at www.sciencemag.org/feature/data/1044449.shl.

All pharmacogenetic polymorphisms studied to date differ in frequency among ethnic and racial groups. In fact, the slow acetylator phenotype was originally suspected to be genetically determined because of the difference in frequency of isoniazid-induced neuropathies observed in Japan versus those observed in the United States (9). The marked racial and ethnic diversity in the frequency of functional polymorphisms in drug- and xenobiotic-metabolizing enzymes dictates that race be considered in studies aimed at discovering whether specific genotypes or phenotypes are associated with disease risk or drug toxicity.

It is now well recognized that adverse drug reactions may be caused by specific drug-metabolizer phenotypes. This is illustrated by the severe and potentially fatal hematopoietic toxicity that occurs when thiopurine methyltransferase-deficient patients are treated with standard doses of azathioprine or mercaptopurine (6). Another example is the slow acetylator phenotype that has been associated with hydralazine-induced lupus, isoniazid-induced neuropathies, dye-associated bladder cancer, and sulfonamide-induced hypersensitivity reactions (9, 10); in all cases, acetylation of a parent drug or an active metabolite is an inactivating pathway. N-Acetyltransferase is an enzyme that conju-

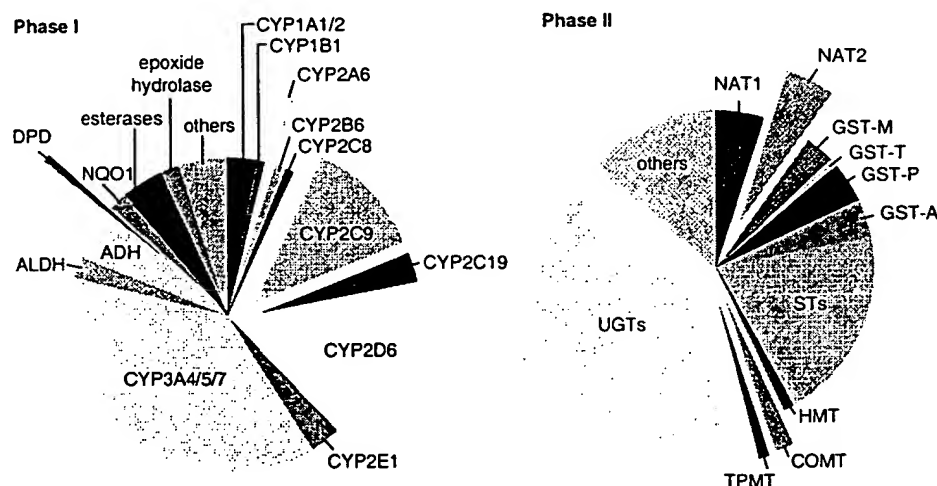


Fig. 2. Most drug-metabolizing enzymes exhibit clinically relevant genetic polymorphisms. Essentially all of the major human enzymes responsible for modification of functional groups [classified as phase I reactions (left)] or conjugation with endogenous substituents [classified as phase II reactions (right)] exhibit common polymorphisms at the genomic level; those enzyme polymorphisms that have already been associated with changes in drug effects are separated from the corresponding pie charts. The percentage of phase I and phase II metabolism of drugs that each enzyme contributes is estimated by the relative size of each section of the corresponding chart. ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; CYP, cytochrome P450; DPD, dihydropyrimidine dehydrogenase; NQO1, NADPH:quinone oxidoreductase or DT diaphorase; COMT, catechol O-methyltransferase; GST, glutathione S-transferase; HMT, histamine methyltransferase; NAT, N-acetyltransferase; STs, sulfotransferases; TPMT, thiopurine methyltransferase; UGTs, uridine 5'-triphosphate glucuronosyltransferases.

case, in the metabolism and disposition of medications, and in the targets of drug therapy. Such diagnostics, which need be performed only once for each battery of genes tested, can then become the blueprint for individualizing drug therapy. This is illustrated in Fig. 3, which depicts various genes that could be genotyped to guide the selection and dosing of chemotherapy for a patient with acute lymphoblastic leukemia (ALL). It is already known that genetic polymorphisms in drug-metabolizing enzymes can have a profound effect on toxicity and efficacy of medications used to treat ALL (6) and that individualizing drug dosages can improve clinical outcome (30). It has also been established that the genotype of leukemic lymphoblasts is an important prognostic variable that can be used to guide the intensity of treatment (31). Furthermore, genetic polymorphisms are also known to exist for cytokines and other determinants of host susceptibility to pathogens, and polymorphisms in cardiovascular, endocrine, and other receptors may be important determinants of an individual's susceptibility to drug toxicity. Putting all of these molecular diagnostics on an "ALL chip" would provide the basis for rapidly and objectively selecting therapy for each patient. These examples represent our current, relatively poor, understanding of genetic determinants of leukemia therapy and host sensitivity to treatment; ongoing studies will provide important insights that should substantially enhance the utility of such pharmacogenomic strategies for ALL and many other human illnesses.

References and Notes

1. W. Kalow, *Lancet* 211, 576 (1956).
2. P. E. Carson, C. L. Flanagan, C. E. Ickes, A. S. Alving, *Science* 124, 484 (1956).
3. H. B. Hughes, J. P. Biehl, A. P. Jones, L. H. Schmidt, *Am. Rev. Tuberc.* 70, 266 (1954); D. A. P. Evans, K. A. Manley, V. A. McKusick, *Br. Med. J.* 2, 485 (1960).
4. D. W. Nebert, *Am. J. Hum. Genet.* 60, 265 (1997); U. A. Meyer and U. M. Zanger, *Annu. Rev. Pharmacol. Toxicol.* 37, 269 (1997).
5. F. J. Gonzalez et al., *Nature* 331, 442 (1988).
6. E. Y. Krynetski and W. E. Evans, *Am. J. Hum. Genet.* 63, 11 (1998).
7. J. Desmeules, M. P. Gascon, P. Dayer, M. Magistris, *Eur. J. Clin. Pharmacol.* 41, 23 (1991); L. Poulsen et al., *ibid.* 51, 289 (1996).
8. A. Mahgoub, J. R. Idle, L. G. Dring, R. Lancaster, R. L. Smith, *Lancet* 2, 584 (1977).
9. D. M. Grant et al., *Mutat. Res.* 376, 61 (1997); S. P. Spielberg, *J. Pharmacokinetic. Biopharm.* 24, 509 (1996).
10. H. Nakamura et al., *J. Pharmacol. Exp. Ther.* 274, 1099 (1995); M. Blum, A. Demierre, D. M. Grant, M. Heim, U. A. Meyer, *Proc. Natl. Acad. Sci. U.S.A.* 88, 5237 (1991).
11. M. J. Ratain et al., *Pharmacogenetics* 6, 93 (1996).
12. S. H. Sindrup, L. Poulsen, K. Brosen, L. Arendt-Nielsen, L. F. Gram, *Pain* 53, 335 (1993).
13. R. B. Diasio, T. L. Beavers, J. T. Carpenter, *J. Clin. Invest.* 81, 47 (1988); F. J. Gonzalez and P. Fernandez-Salguero, *Trends. Pharmacol. Sci.* 16, 325 (1995).
14. T. R. Rebbeck, J. M. Jaffe, A. H. Walker, A. J. Wein, S. B. Malkowicz, *J. Natl. Cancer Inst.* 90, 1225 (1998); C. A. Felix et al., *Proc. Natl. Acad. Sci. U.S.A.* 95, 13176 (1998).
15. E. G. Schuetz and A. H. Schinkel, *J. Biochem. Mol. Toxicol.* 13, 219 (1999); E. G. Schuetz, W. T. Beck, J. D. Schuetz, *Mol. Pharmacol.* 49, 311 (1996); V. J. Wachter, C. Y. Wu, L. Z. Benet, *Mol. Carcinog.* 13, 129 (1995).
16. A. H. Schinkel, *Semin. Cancer Biol.* 8, 161 (1997); L. A. Mickley et al., *Blood* 91, 1749 (1998); N. Kioka et al., *Biochem. Biophys. Res. Commun.* 162, 224 (1989).
17. F. D. Martinez et al., *J. Clin. Invest.* 100, 3184 (1997); R. J. Hancox, M. R. Sears, D. R. Taylor, *Eur. Respir. J.* 11, 589 (1999); B. J. Lipworth, I. P. Hall, S. Tan, I. Aziz, J. C. Crabbe, *Chest* 115, 324 (1999); J. J. Lima et al., *Clin. Pharmacol. Ther.* 65, 519 (1999).
18. F. G. van der Kleij et al., *Nephrol. Dial. Transplant.* 12 (suppl. 2), 42 (1997); M. Haas et al., *Kidney Blood Press. Res.* 21, 66 (1998); F. G. van der Kleij et al., *Kidney Int. Suppl.* 63, S23 (1997); Y. Nakano et al., *Am. J. Hypertens.* 10, 1064 (1997); L. O'Toole, M. Stewart, P. Padfield, K. Channer, *J. Cardiovasc. Pharmacol.* 32, 988 (1998); M. Sasaki et al., *J. Hypertens.* 14, 1403 (1996); S. Mizuiri et al., *Nephron* 75, 310 (1997); H. Yoshida et al., *J. Clin. Invest.* 96, 2162 (1995).
19. D. Henrion et al., *J. Vasc. Res.* 35, 356 (1998).
20. A. Benetos et al., *Hypertension* 28, 1081 (1996).
21. C. G. V. Essen et al., *Lancet* 347, 94 (1996).
22. M. Arranz et al., *ibid.* 346, 281 (1995).
23. J. Poirier, *Proc. Natl. Acad. Sci. U.S.A.* 92, 12260 (1995).
24. J. A. Kuivenhoven, *N. Engl. J. Med.* 338, 86 (1998).
25. V. M. Steen, R. Lovlie, T. MacEwan, R. G. McCreadie, *Mol. Psychiatry* 2, 139 (1997); C. H. Chen, F. C. Wei, F. J. Koong, K. J. Hsiao, *Biol. Psychiatry* 41, 827 (1997).
26. G. W. Abbott et al., *Cell* 97, 175 (1999).
27. E. F. Gillard et al., *Genomics* 13, 1247 (1992).
28. J. Cohen, *Science* 275, 776 (1997); A. Marshall, *Nature Biotechnol.* 15, 954 (1997).
29. F. S. Collins, *N. Engl. J. Med.* 341, 28 (1999); P. W. Kleyn and E. S. Vesell, *Science* 281, 1820 (1998).
30. W. E. Evans et al., *N. Engl. J. Med.* 338, 499 (1998).
31. C. H. Pui and W. E. Evans, *ibid.* 339, 605 (1998).
32. G. P. Aithal, C. P. Day, P. J. Kesteven, A. K. Daly, *Lancet* 353, 717 (1999); D. J. Steward et al., *Pharmacogenetics* 7, 361 (1997).
33. T. Kapitany et al., *Schizophren. Res.* 32, 101 (1998); Y. Caraco, J. Sheller, A. J. Wood, *J. Pharmacol. Exp. Ther.* 278, 1165 (1996); R. F. Tyndale, K. P. Droll, E. M. Sellers, *Pharmacogenetics* 7, 375 (1997); K. Brosen, *Eur. J. Clin. Pharmacol.* 36, 537 (1989); H. H. Zhou, R. P. Koshakji, D. J. Silberstein, C. R. Wilkinson, A. J. Wood, *N. Engl. J. Med.* 320, 565 (1989); M. S. Lennard et al., *Clin. Pharmacol. Ther.* 34, 732 (1983); J. T. Lee et al., *N. Engl. J. Med.* 322, 1764 (1990).
34. M. V. Relling, M. L. Hancock, J. M. Boyett, C.-H. Pui, W. E. Evans, *Blood* 93, 2817 (1999); M. V. Relling et al., *Lancet* 354, 34 (1999); A. J. Black et al., *Ann. Intern. Med.* 129, 716 (1998); L. Lennard, *Ther. Drug Monit.* 20, 527 (1998); J. Aarbakke, G. Janka-Schaub, G. B. Elion, *Trends. Pharmacol. Sci.* 18, 3 (1997); L. Lennard, J. A. Van Loon, J. S. Lilleyman, R. M. Weinshilboum, *Clin. Pharmacol. Ther.* 41, 18 (1987).
35. S. G. Priori et al., *Circulation* 99, 674 (1999); C. Donger et al., *ibid.* 96, 2778 (1997).
36. Supported in part by grants NIH R37 CA36401, RO1 CA51001, and RO1 CA78224; Cancer Center support grant CA21765; a Center of Excellence grant from the state of Tennessee; and the American Lebanese Syrian Associated Charities (ALSAC). We thank J. Johnson, M. Kastan, C. Naev, D. Roden, E. Schuetz, J. Schuetz, and E. Tuomanen for helpful suggestions; L. Rawlinson for artistic contributions to Figs. 1 and 3; and S. Davis, G. McDugle, and F. Munn for help in preparing the manuscript.

So instant, you don't need water...

NEW! Science Online's Content Alert Service: The only source for instant updates on breaking science news. This free Science Online enhancement e-mails summaries of the latest research articles published weekly in *Science* – **instantly**. To sign up for the Content Alert service, go to Science Online – and save the water for your coffee.

Science
www.sciencemag.org

For more information about Content Alerts go to www.sciencemag.org. Click on Subscription button, then click on Content Alert button

Clp

RESEARCH

A Novel In Vivo Method to Detect DNA Sequence Variation

Malek Faham¹ and David R. Cox^{2,3}

¹Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143; ²Department of Genetics and Pediatrics, Stanford University, School of Medicine, Stanford, California 94305

Mismatch repair detection (MRD) is an in vivo method that uses a change in bacterial colony color to detect DNA sequence variation. DNA fragments to be screened for variation are cloned into two MRD plasmids, and bacteria are transformed with heteroduplexes of these constructs. The resulting colonies are blue in the absence of a mismatch and white in the presence of a mismatch. MRD is capable of detecting a single mismatch in a DNA fragment as large as 10 kb in size. In addition, MRD has the potential for analyzing many fragments simultaneously, offering a powerful method for high-throughput genotyping and mutation detection in a large genomic region.

The detection of mutations in genomic DNA plays a critical role in efforts to elucidate the genetic basis of human disease. Although many approaches are currently applied to the problem of mutation detection, no single technique provides a rapid method for screening large stretches of genomic DNA with high sensitivity and specificity (Grompe 1993). We have developed an in vivo bacterial assay, mismatch repair detection (MRD), that utilizes the *Escherichia coli* methyl-directed mismatch repair system to detect single-base mismatches in DNA. Unlike other DNA variation detection techniques, MRD can detect a single-base mismatch in up to 10 kb of DNA. In addition, MRD has the potential to examine many different DNA fragments simultaneously, providing a rapid method for screening large stretches of DNA for nucleotide sequence variation.

The normal function of the *E. coli* methyl-directed mismatch repair system is to correct errors in newly synthesized DNA resulting from imperfect DNA replication (Wagner and Meselson 1976). The system distinguishes unreplicated from newly replicated DNA by taking advantage of the fact that methylation of adenine in the sequence GATC occurs in unreplicated DNA but not in newly replicated DNA. Mismatch repair is initiated by the action of three proteins, MutS, MutL, and MutH, which lead to nicking of the

unmethylated, newly replicated DNA strand at a hemimethylated GATC site. The unmethylated DNA strand is then digested and resynthesized in a replication reaction in which the methylated strand is used as a template (Modrich 1991). The methyl-directed mismatch repair system can repair single-base mismatches and loops up to 3 nucleotides in length. Loops of 5 nucleotides and larger are not repaired (Parker and Marinus 1992). We have taken advantage of the inability of the mismatch repair system to repair loops of 5 nucleotides or greater to design two vectors that allow in vivo mismatch repair to be detected visually as a change in bacterial colony color.

Two pUC-derived plasmids, the blue (pMF200) and the white (pMF100) plasmid, are used in the MRD procedure. These plasmids are identical except for a 5-bp insertion into the *lacZα* gene of pMF100 (Fig. 1). This insertion results in white colonies when bacteria transformed with the plasmid are grown on LB plates supplemented with indolyl-β-D-galactoside (X-gal) and isopropyl-β-D-thiogalactoside (IPTG). In contrast, bacteria transformed with the blue plasmid result in blue colonies when grown under these conditions. The initial step of the MRD procedure (Fig. 2) consists of cloning one of two DNA fragments to be screened for differences into the blue plasmid and cloning of the second DNA fragment into the white plasmid. The blue plasmid construct is then transformed into a *dam*⁻ bacterial strain, resulting in a completely unmethylated plasmid, whereas the white plas-

³Corresponding author.
E-MAIL cox@shgc.stanford.edu; FAX (415) 725-8058.

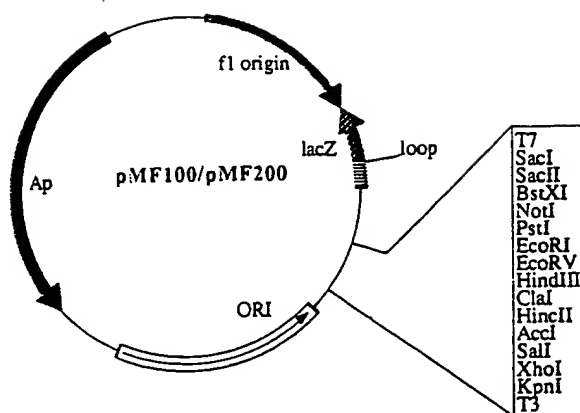


Figure 1 The MRD vectors. pMF100 and pMF200 are derived from pUC19 with the multiple cloning site displaced from the *lacZ* region (Yanisch-Perron et al. 1985). In addition, the MRD vectors contain the *Bgl*I fragment (2166–472) and most of the multiple cloning site of pBluescript (Short et al. 1988). The multiple cloning sites of the MRD vectors do not have sites for the restriction enzymes *Xba*I, *Spe*I, *Bam*HI, *Sma*I, and *Apa*I; the *Eco*RI site is not unique. pUC19 multiple cloning sites (nucleotides 400–454) were replaced using 70-nucleotide-long oligonucleotides with a sequence containing 4 GATC sites. In addition, the sequence replacing the pUC19 multiple cloning sites in pMF200 has a 5-bp insertion as compared to pMF100 creating a nonfunctional *lacZ* in pMF200. The label loop is to indicate this difference between pMF100 and pMF200.

mid construct is transformed into a *dam*⁺ bacterial strain, resulting in a fully methylated plasmid. The two plasmids are then linearized, denatured, and reannealed, resulting in two heteroduplex and two homoduplex plasmids. Following digestion with *Mbo*I, which digests only unmethylated homoduplexes, and *Dpn*I, which digests only fully methylated homoduplexes, the remaining hemimethylated heteroduplexes are circularized, transformed into *E. coli*, and plated onto agar supplemented with X-gal and IPTG. In the absence of a mismatch between the two test DNA fragments, the 5-nucleotide loop in the *lacZ* gene that results from heteroduplex formation between the white and the blue plasmids is not repaired by the mismatch repair system. Subsequent plasmid replication produces both white and blue plasmids in a single colony, leading to a blue color. In contrast, if a mismatch is present in the heteroduplex DNA, a corepair event takes place that involves both the mismatch in the DNA as well as the 5-nucleotide loop in the *lacZ* gene. In this case, the

IN VIVO DETECTION OF DNA SEQUENCE VARIATION

unmethylated *lacZ* gene on the blue plasmid is degraded and replaced by the *lacZ* gene from the methylated strand of the white plasmid, resulting in a white colony. Previous *in vivo* studies have suggested that the corepaired segment of DNA is at least 1.5 kb (Carraway and Marinus 1993). We have found that corepair of a mismatch and the *lacZ* gene in the MRD system occurs even when the distance between them is 5 kb (see below).

RESULTS

Testing Known Mutations

As an initial test of the sensitivity and specificity of the MRD system, we tested the detection of a single-nucleotide mismatch in a 550-bp DNA fragment derived from the promoter of the mouse β -globin gene (Myers et al. 1985a). We used MRD to compare this DNA fragment, which contains a T at position –49 relative to the functional transcription start site of the gene, with a second DNA fragment identical in sequence except for a C at position –49. In this experiment, the mismatch was located ~700 bp from the 5-nucleotide *lacZ* loop in the vector. Comparison of the two DNA molecules by using MRD resulted in 90% white colonies. In contrast, comparison of the same two DNA molecules with no mismatch (–49T/–49T) resulted in only 7% white colonies (Table 1; Figs. 2 and 3). Comparison of all of the possible different single-nucleotide mismatches at position –49 using MRD revealed proportions of white colonies ranging from 80% to 90% (Table 1; Figs. 2 and 3). These results demonstrate that MRD can detect all of the different DNA variations possible at this position with high efficiency.

In an effort to establish the generality of the above results, we used the MRD system to detect a total of five additional single-nucleotide mismatches in two different DNA fragments (Table 1). Four of these mismatches are at different nucleotide positions in the human cystathionine β -synthase gene (Kruger and Cox 1995). The remaining one mismatch represents a single-nucleotide change in the human agouti gene (Wilson et al. 1995). In each case, we were able to detect the single-nucleotide mismatch (Table 1).

The detection of a single mismatch in 10 kb of DNA

In the experiments described above, we were surprised to observe that we were able to detect the

mismatch even when it was as far from the loop as 2.3 kb. In addition, because the proportion of white colonies was >50%, corepair of the mismatch and the loop occurred irrespective of which side the mismatch was located relative to the loop on the unmethylated strand. In an effort to determine whether the efficiency of mismatch detection would remain high if the distance between a mismatch and the vector loop was even larger, we performed the following experiment. A 9-kb test DNA fragment derived from bacteriophage λ was cloned into the MRD plasmid system and compared with the same test DNA containing a 2-bp insertion located 5 kb from one end of the fragment. Because DNA molecules used for transformation are circular, a mismatch in a 10-kb fragment is always within 5 kb of the loop. The mismatch in this experiment was at least 5 kb away from the loop in either direction. In the presence of the 2-bp loop, 70% white colonies were produced, as compared with 10% white colonies in the absence of the mismatch. These results indicate that MRD can detect a mismatch in 10 kb of DNA.

Detecting variation in PCR products

Next, we investigated the utility of MRD for detecting unknown mutations in genomic DNA fragments generated by the polymerase chain reaction (PCR). PCR is a practical method for obtaining a particular genomic DNA fragment of interest from many different individuals. Recent advances in PCR technology make it possible to isolate DNA products >10 kb in length (Barnes 1994; Cheng et al. 1994). However, the introduction of errors during the PCR reaction severely limits the use of individual cloned PCR products for mutation detection, particularly in the case of long PCR products. In an effort to overcome this limitation, we have developed a protocol that uses MRD to enrich for molecules that are free of PCR-

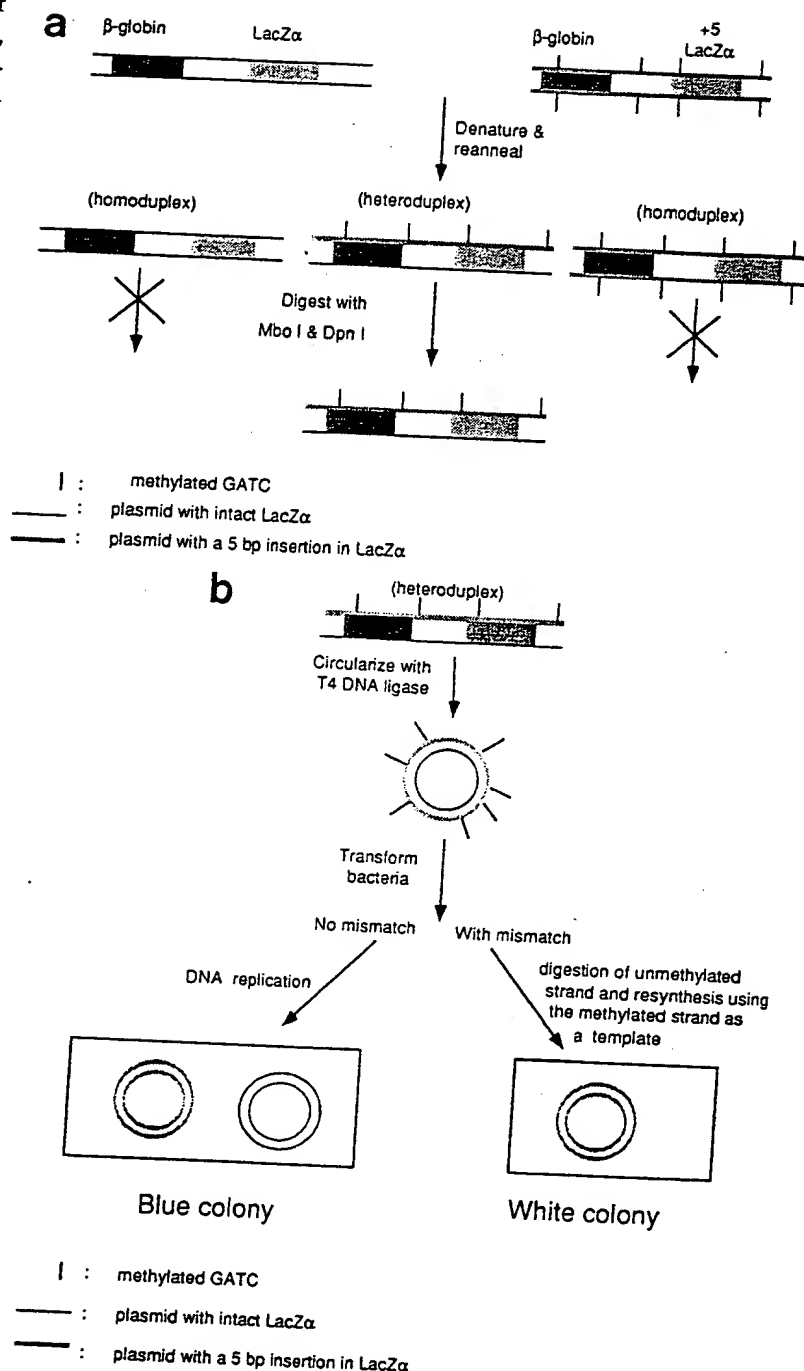


Figure 2 (See facing page for legend.)

induced errors. Following this "cleaning" protocol, the cloned PCR products can be compared for DNA sequence differences by using the MRD procedure described above.

The basic principle underlying the MRD cleaning protocol is the fact that any single PCR-induced mutation makes up a very small fraction

Table 1. Detection of Known Point Mutations Using MRD

Variation ^a	Fragment size ^b	Distance from loop ^b	Percent white colonies ^c
None ^d	0.55	N.A.	7
G/C ^d	0.55	0.7	89
A/T ^d	0.55	0.7	84
G/T ^d	0.55	0.7	82
A/C ^d	0.55	0.7	82
C/T ^d	0.55	0.7	90
None ^e	2.0	N.A.	8
A/C ^e	2.0	0.4	35
None ^f	2.2	N.A.	10
C/T ^f	2.2	2.3	83
G/A ^f	2.2	2.1	86
C/T ^f	2.2	1.6	81
T/C ^f	2.2	1.8	80

The mutations are in the order listed: C341T, G502A, C992T, and T833C.

^a(A/T) At the only position of variation between the two fragments compared, the *dam*⁻-grown variant has an A and the *dam*⁺-grown variant has a T at the same position on the same strand. Therefore, mismatches produced in such an experiment are A/A and T/T.

^bIn kilobases. (N.A.) Not applicable.

^cAt least 250 colonies were counted to determine the percentage.

^dExperiment using fragment of the mouse β -globin gene.

^eExperiment using fragment of the human agouti gene.

^fExperiment using fragment of the human cystathionine β -synthase gene.

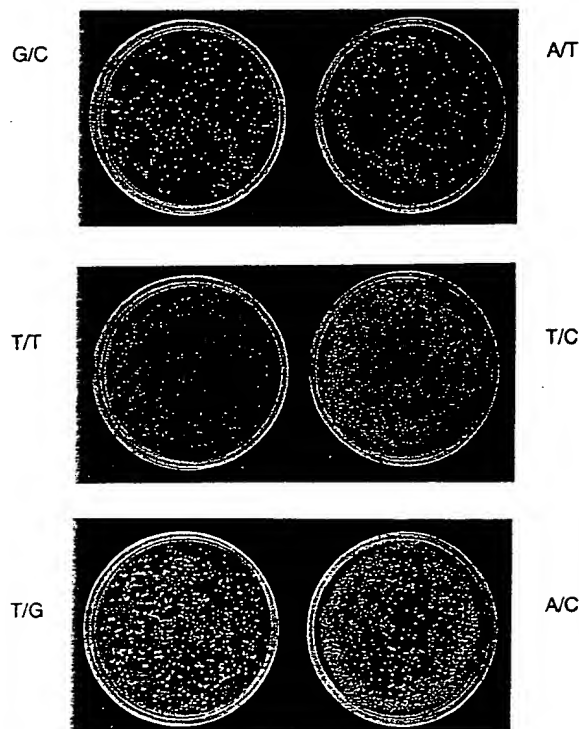


Figure 3 Transformation plates of the different mismatches at position -49 of mouse β -globin promoter. The plate labeled T/T, containing a majority of blue colonies, represents the transformation with the nonmismatched control, the remaining plates, containing a majority of white colonies, represent transformation with mismatched molecules. Nomenclature of the different comparisons is as described in Table 1.

of all the molecules generated by PCR. As a result, when the products of a PCR reaction are cloned into the blue and the white MRD vectors and assayed as described above, the majority of products containing PCR-induced errors are present as heteroduplex molecules containing a mismatch and produce white colonies. In contrast, those PCR products with no PCR-induced errors con-

tain no mismatches and result in blue colonies. Given that not all mismatches are repaired with 100% efficiency, some blue colonies can be expected to contain PCR-induced errors following the first round of enrichment. However, if blue colonies are isolated and used in a second round of MRD cleaning, those molecules containing PCR-induced errors can be reduced even further.

Figure 2 The MRD procedure. (a) Formation of the heteroduplex. DNA from the unmethylated blue plasmid and the methylated white plasmid containing the fragments to be screened are linearized, denatured, and reannealed. The resulting molecules are fully unmethylated blue plasmid homoduplex, fully methylated white plasmid homoduplex, and hemimethylated heteroduplexes (two populations of heteroduplexes are formed). Only the heteroduplex molecules are left intact after treatment with *Mbo*I, which digests fully unmethylated DNA, and *Dpn*I, which digests fully methylated DNA. (b) Introduction of the heteroduplex into *E. coli* and detection of the variation. The heteroduplex molecules prepared in a are circularized with T4 DNA ligase and transformed into *E. coli*. In the absence of a mismatch, DNA replication in the bacteria generates both the blue and the white plasmid, producing a blue colony. In the presence of a mismatch, repair of the unmethylated blue strand of the heteroduplex using the white strand as a template generates the white plasmid only, producing a white colony.

Because each blue colony contains both a blue and a white MRD plasmid (see above), the second round of MRD cleaning is carried out as follows.

Plasmid DNA isolated from blue colonies following the first round of cleaning is used to transform both *dam*⁻ and a *dam*⁺ bacterial strains. Although both blue and white colonies result from each transformation, only the blue colonies are isolated from the *dam*⁻ transformation and only the white colonies are isolated from the *dam*⁺ transformation. Plasmid DNA is prepared from such colonies, and heteroduplexes are isolated as described above. Blue colonies arising from transformation with these heteroduplexes are enriched further for the products free of PCR-induced error. For example, in an experiment in which 75% of molecules contain one or more PCR-induced errors following PCR, assuming 95% efficiency of mismatch repair and 10% frequency of white colonies in the absence of a mismatch, the expectation would be 10% blue colonies following one round of MRD enrichment, with 66% of the molecules in such colonies free of PCR-induced errors. If the plasmid DNA from the blue colonies were used for a second round of MRD enrichment, the expectation would be 41% blue colonies, with 96% of the molecules in such colonies free of PCR-induced errors.

As a test of the practicality and the efficiency of the MRD cleaning protocol, we isolated a 2-kb human chromosome 21-specific PCR product from each of the two chromosome 21 homologs of a single individual. The two chromosome 21 homologs were separated from each other in independent hamster-human somatic cell hybrid clones. Genomic DNA isolated from these somatic cell hybrid clones was the template of the PCR reactions. When the PCR products derived from each homolog were compared by using MRD as described above, ~10% blue colonies were observed in each case. Following two rounds of MRD cleaning, the proportion of blue colonies was 60%–80% (Table 2). In contrast, when these "cleaned" PCR products derived from the two homologs were compared with each other by using MRD, ~90% of the resulting colonies were white, indicating the presence of at least one single-base difference in the 2-kb PCR products derived from the two different chromosome 21 homologs. We have demonstrated independently the presence of at least one DNA sequence variation in these 2-kb PCR products by finding a *Hinf*I restriction fragment length polymorphism (RFLP) (data not shown). These results

Table 2. Detection of DNA Variation Following PCR Cleaning

Variants compared ^a	Experimental stage	Percent white colonies ^b
1/1	no cleaning	>90
2/2	no cleaning	>90
B1/B1	one cleaning round	70
B2/B2	one cleaning round	64
BlueB1/BlueB1	two cleaning rounds	38
BlueB2/BlueB2	two cleaning rounds	21
BlueB1/BlueB2	testing clean products	>90
BlueB2/BlueB1	testing clean products	>90

^a1 and 2 represent PCR products from two different homologs of chromosome 21 isolated in hamster-human somatic cell hybrids. (1/1) Comparison of a blue plasmid grown in a *dam*⁻ strain containing the PCR product from hybrid 1 with a white plasmid grown in a *dam*⁺ strain containing the PCR product from hybrid 1. (B1/B1) Comparison of blue *dam*⁻ grown plasmids obtained from the blue colonies of comparison 1/1 to white *dam*⁺-grown plasmids obtained from the same source. (BlueB1/BlueB1) Comparison of blue *dam*⁻ grown plasmids obtained from the blue colonies of comparison B1/B1 to white *dam*⁺-grown plasmids from the same source. (BlueB1/BlueB2) Comparison of blue *dam*⁻-grown plasmids from the blue colonies of comparison B1/B1 to white *dam*⁺-grown plasmids obtained from the blue colonies of comparison B2/B2.

^bAt least 200 colonies were counted to determine the percentage.

demonstrate that MRD can be used to enrich for PCR products that are largely free of PCR-induced errors and that such products can be used in conjunction with MRD to detect human DNA sequence variation. In addition, we have used MRD in conjugation with the high-fidelity polymerase, *Pfu*, to analyze 2-kb PCR products for DNA variations without the need to perform PCR cleaning (M. Faham and D.R. Cox, unpubl.).

DISCUSSION

Current techniques for detecting unknown mutations in genomic DNA fall into three general classes. The first class of techniques, which in-

cludes single-strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices (Myers et al. 1985b,c; Orita et al. 1989a,b; Sheffield et al. 1989; Perry and Carrell 1992; White et al. 1992), detects conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. These techniques are limited by the need to determine optimum reaction conditions for each DNA fragment and by a marked decrease in sensitivity with increasing DNA fragment size. The second class of techniques, which includes RNase A cleavage, chemical mismatch cleavage (CMC), and enzyme mismatch cleavage (EMC), uses chemicals or proteins to detect sites of sequence mismatch in heteroduplex DNA (Myers et al. 1985d; Cotton et al. 1988; Maschal et al. 1995; Youil et al. 1995). These techniques can be used to assay many different DNA fragments with a single set of assay conditions. In addition, they can be used to detect mutations in larger DNA fragments. However, even with this second class of techniques, the upper limit for the size of the screened DNA fragment is ~1 kb.

Unlike all of the techniques described above, which involve in vitro analyses of DNA to detect sequence variation, MRD utilizes an in vivo assay for detecting unknown mutations in genomic DNA. We have used this system to analyze a variety of heteroduplex molecules with inserts ranging in size from 550 to 9 kb, representing each of the four possible classes of single-nucleotide substitutions between the strands. All of the mutations tested, nine of nine point mutations (Table 1) and three of three deletions of 2–3 bp (M. Faham and D.R. Cox, unpubl.), could be detected unambiguously. Our data indicate that MRD can detect mismatches in DNA fragments of up to 10 kb in size. Thus, MRD overcomes one of the major limitations of techniques currently available for detecting unknown mutations.

In some cases of mutation detection (e.g., comparison of a patient's DNA with the patient's tumor DNA) and polymorphism detection (e.g., identification of a polymorphic marker for mapping a recombination breakpoint), the goal of the experiment is to identify a variant DNA fragment. In such cases, MRD's ability to detect DNA variation in long DNA fragments with high sensitivity is particularly useful. In other cases of mutation detection in human genomic DNA or high throughput genotyping, the experimental goal is

to identify which one of many variations in a large genomic region is the disease-causing variation. To achieve this goal, one needs to test for the presence of the different identified variations in many people from the normal population. In such cases, an efficient analysis of many small fragments is more beneficial than the analysis of a long DNA fragment. MRD is well suited for this experimental problem, as the technique can be used to analyze many fragments simultaneously in a single experiment. In such an experiment, heteroduplexes are made between a pool of restriction fragments from the genomic region of interest of a "standard" and a "test". This is followed by ligation of these heteroduplexes into the hemimethylated MRD heteroduplex vector and transformation into *E. coli*. The resulting blue colonies contain DNA fragments that have no sequence variation between the tester and the standard, whereas the white colonies contain fragments with sequence differences. DNA prepared from the pool of blue colonies contains fragments of identity, whereas DNA prepared from the pool of white colonies contains fragments containing differences. Determination of whether a specific DNA fragment is present in the white pool or the blue pool indicates whether the fragment contains a variation. One can use agarose gel electrophoresis of a restriction digest that releases the insert fragments in the blue and the white pools to determine the pool in which each fragment is present. We performed this procedure to analyze up to 10 DNA fragments simultaneously for variation (M. Faham and D.R. Cox, unpubl.). To analyze more fragments, the resolution of the different fragments on agarose gels would be impractical. One can use the blue and white pool DNA as independent hybridization probes on a blot containing DNA from each of the different fragments. For each dot, the comparison of the hybridization signal produced by the blue probe with that produced by the white probe determines whether that fragment contains a variation. Such a procedure has the potential for detecting the presence of mutations in a region representing hundreds of kilobases of DNA or for genotyping many loci simultaneously. This approach is similar to that used in the genomic mismatch scanning (GMS) procedure for identifying regions of the genome identical by descent (Nelson et al. 1993). However, an important difference between GMS and MRD is that GMS yields a probe only for regions of identity, whereas MRD yields probes for both regions

of identity and regions of difference. The ability to isolate both types of probes results in improved signal to noise as compared to the use of a single probe. Although the manual isolation of blue from white colonies becomes impractical as the number of colonies becomes large, the use of an automatic cell sorter or a selectable system to isolate blue from white colonies should allow for the analysis of very large numbers of colonies.

The detection of DNA variation by MRD is limited by the ability to obtain the specific DNA fragments to be analyzed from the patients of interest. However, several approaches are presently available to isolate the necessary DNA fragments, including long-range PCR with high-fidelity enzymes (e.g., *Pfu*) (Neilson et al. 1995), *recA*-assisted cleavage (Ferrin and Camerini-Otero 1991), and the use of a single set of oligonucleotide primers to amplify multiple specific fragments simultaneously by PCR (Brookes et al. 1994). In conjunction with these methods for isolating specific genomic DNA fragments, MRD provides a powerful technique for the detection of unknown mutations, the detection of DNA variation in large genomic regions, and high-throughput genotyping.

METHODS

Construction of MRD Vectors

pNEB193 (New England Biolabs) was digested with *AseI*, and the larger two fragments were ligated to each other leading to a construct, pNEB133, with a deletion of ~60 bp. pBCKS (Stratagene) was digested with *SpeI*, filled in with Klenow, digested with *SmaI*, and recircularized with T4 DNA ligase. The resultant clone ΔB was resistant to cleavage by *XbaI* in addition to the other expected enzymes (*SpeI*, *BamHI*, and *SmaI*). *BssHII* digest of ΔB was performed, and the smaller band (~150 bp) was gel eluted and ligated to *TfiI*-digested and Klenow-filled pNEB133. The resultant clone, pNEB133B, had the orientation of T3 being the far end from the *lacZ α* gene. The *BspHI* fragment containing the ampicillin gene of pNEB133B was replaced with a PCR-generated fragment carrying the chloramphenicol-resistant gene producing pMF0. pNEB133B was digested with *ApaI*, the 3' overhang chewed with T4 DNA polymerase, a *BglII* linker added, and recircularization performed with T4 DNA ligase producing the clone pNEB133BB. The *BglII* fragment containing the M13 origin of replication from pBluescript was inserted instead of the *BglII* fragment of pNEB133BB, producing the clone pBBM. Two complementary oligonucleotides were cloned into an *EcoRI*-*HindIII*-digested pNEB133. One oligonucleotide had the following sequence: 5'-AATTCTGCACGGATCCACGC-GATCGCTCTGATCAGCAGATCTCACTG GTGACCTCT-TAATTAACAGCATGC-3'. The other oligonucleotide had the complementary sequence except it had the sequence AGCT as an overhang at its 5' end and it lacked a complementary to the last 4 nucleotides of the 5' end of the first

oligonucleotide. The resultant clone, pOI was digested with *BclI* and *EcoRI* and ligated to 2 complementary oligonucleotides that are identical to the sequence deleted in pOI except that they lack the 5 nucleotides GCACG, destroying the *BamHI* site and producing a clone, pOII, that has a deletion just upstream of the *EcoRI* site that makes *lacZ α* in-frame. The *A/III*-*BglII* fragment of pMF0 was replaced with the *A/III*-*BglII* fragments of pOI and pOII producing two chloramphenicol-resistant clones that are identical except for a 5-bp insertion. The two clones were named pMF2 and pMF2-5; pMF2 produces white colonies in the proper medium and is the product of pOI, and pMF2-5 produces blue colonies in the proper medium and is the product of pOII. The smaller *EcoRI* fragment of pBBM was replaced with the small *EcoRI* fragments of pMF2 and pMF2-5 producing two ampicillin-resistant clones that are exactly identical except for a 5-bp insertion.

Testing Known Mutations

A 550-bp *Clal*-*SacI* fragment of four variants having A, G, T, or C at position -49 of mouse β -globin promoter, was cloned into *Clal*-*SacI* pMF100 and pMF200. The T and C variants were cloned into pMF100; the T, G, and A were cloned into pMF200. pMF200 clones were grown in a *dam*⁻ strain (SCS110) (Stratagene), and pMF100 clones were grown in a *dam*⁺ strain (DH5 α) (GIBCO-BRL). The plasmid DNA was linearized with *A/III* in 30- μ l reactions. About equal amounts (estimated by gel electrophoresis of the linearized plasmids) were mixed, and the volume increased to 100 μ l with TE buffer (10 mM Tris, 0.1 mM EDTA). The sample was then extracted with 100 μ l of a 1:1 phenol/chloroform mixture, followed by extraction with 100 μ l of chloroform. Five microliters of 0.5 M EDTA and 12.5 μ l of 1 M NaOH were added, and the reaction left at room temperature for 15 min. The reaction was neutralized by the addition of 12.5 μ l of 2 M Tris (pH 7.2), 125 μ l of formamide was added, and the reaction incubated at 30°C for 1 hr. Chloroform extraction was performed twice, followed by ethanol precipitation. DNA was digested in a 20- μ l reaction with *MboI* (5 units) for 1 hr, and *DpnI* (10–20 units) for 10 min at 37°C. The reaction was stopped by the addition of 1 μ l of 0.5 mM EDTA. The intact heteroduplex plasmid was separated from the *MboI*- or *DpnI*-digested plasmid by agarose gel electrophoresis. The DNA was isolated from the gel slice and resuspended in 20 μ l of water. Three microliters was used for a 20- μ l recircularization reaction using T4 DNA ligase overnight at 16°C. Transformation of DH5 α was performed with only 30 min recovery at 37°C after the heat shock. Transformation reactions were plated on LB agar plates with 50 μ g/ml of carbinicillin, 64 μ g/ml of X-gal, and 64 μ g/ml of IPTG and incubated at 37°C overnight. *EcoRI*-*PvuII* fragments of the cystathionine β -synthase alleles were obtained from constructs of Kruger and Cox (1995). Adapters converting the *EcoRI* overhang to *NotI* overhangs were ligated on the fragments. These fragments were subsequently cloned in a *NotI*-*EcoRV*-digested pMF100. In addition, the wild-type control fragment was cloned in the pMF100. The MRD analysis was performed as in above. The control experiment was comparing the blue vector (pMF200) carrying the wild-type allele to the white vector (pMF100) carrying the same allele. The test experiment compared the blue vector carrying the wild-type allele to the white vector carrying another allele.

KpnI-*EcoRV* fragments of human agouti alleles were obtained from constructs of Wilson et al. (1995). These fragments were cloned in *KpnI*-*EcoRV*-digested MRD vectors. The wild-type allele was cloned in both pMF200 and pMF100. The mutant allele was cloned in pMF100 only. The MRD procedure was performed as described above. The control experiment compared the blue vector carrying the wild-type allele to the white vector carrying the wild-type allele. The test experiment compared the blue vector carrying the wild-type allele to the white vector carrying the mutant allele.

A 9-kb fragment of λ DNA was cloned in pMF1 (a relative vector to pMF2-5) producing the clone pB10. Partial digest with *BstBI* was performed. (Two *BstBI* sites were present: one in the chloramphenicol resistant gene, and the other -5 kb in the insert.) Fill-in reaction was performed with Klenow, followed by recircularization with T4 DNA ligase and transformation into bacteria. Resultant clones were analyzed for the generation of a *NruI* site at the correct position; the correct clone was named pB10 + 2. A *NotI*-*KasI* fragment containing the insert of pB10 was cloned in a *NotI*-*KasI*-digested pMF2. The resulting clone pB210 was compared using MRD to pB10 and pB10 + 2. The MRD procedure was performed as described above.

Detecting Variation in a PCR Product

An *EcoRI* fragment of a cosmid was subcloned in pNEB193. Using sequence information of the clone, primers were designed to produce a PCR product of ~3 kb in size. The PCR reaction was performed with the enzyme mixture *rTth* + *Vent* (Perkin Elmer). The PCR product was extracted with an equal volume of a 1:1 mixture of phenol/chloroform and ethanol precipitated. Restriction digest with *HindIII* and *KpnI* was performed producing a fragment of ~2 kb in size. These fragments were cloned in *HindIII*-*KpnI*-digested vectors that are relatives to pMF100 and pMF200. Only 5% of the transformation mixture was plated, and the rest was grown directly in 5 ml of Luria broth (LB) containing 50 μ g/ml of carbinicillin. DNA was isolated from the transformation cultures and DNA of the pMF200 clones was transformed into SCS110 (*dam*⁻ strain). Five percent of the transformation was plated and the rest grown directly in 5 ml of LB + 50 μ g/ml of carbinicillin. DNA isolated from these cultures was compared with DNA isolated from the pMF100 clones carrying fragments generated from the same source (i.e., the same somatic cell hybrid). The MRD procedure was performed as described above. About 50 blue colonies from this comparison were picked and grown in 5 ml of LB + 50 μ g/ml of carbinicillin. One microliter of a 1:1000 dilution of DNA isolated from these cultures was used to transform DHS α (*dam*⁺ strain), and 2 μ l of the same dilution was used to transform SCS110 (*dam*⁻ strain). White colonies from the first transformation and blue colonies from the second transformation were picked and grown in 5 ml of LB + 50 μ g/ml of carbinicillin. DNA isolated from these cultures was used to perform the MRD procedure. Subsequent to this second round, blue colonies were picked and grown, and their DNA was used to transform the two bacterial strains as described above. These DNA samples were used for comparison fragments generated from the same source and fragments generated from a different source (i.e., the other somatic cell hybrid).

ACKNOWLEDGMENTS

We are grateful to R.M. Myers, G. Barsh, and W. Kruger for generously contributing the mutations in the mouse β -globin, the human agouti, and the human cystathionine β -synthase genes, respectively. We thank the members of the Cox laboratory and C. Iannicola for helpful discussions. This work has been supported by National Institutes of Health grant HD 24610 to D.R.C.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci.* **91**: 2216-2220.
- Brookes, A.J. et al. 1994. Cloning the shared components of complex DNA resources. *Hum. Mol. Genet.* **3**: 2011-2017.
- Carraway, M. and M.G. Marinus. 1993. Repair of heteroduplex DNA molecules with loops in *Escherichia coli*. *J. Bacteriol.* **175**: 3972-3980.
- Cheng, S., C. Fockler, W.M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci.* **91**: 5695-5699.
- Cotton, R.G.H., N.R. Roguere, and R.D. Campbell. 1988. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its applications to the study of mutations. *Proc. Natl. Acad. Sci.* **85**: 4397-4401.
- Ferrin, L.J. and R.D. Camerini-Otero. 1991. Selective cleavage of human DNA: *recA*-assisted restriction endonuclease (RARE) cleavage. *Science* **254**: 1494-1497.
- Grompe, M. 1993. The rapid detection of unknown mutations in nucleic acids. *Nature Genet.* **5**: 111-117.
- Kruger, W.D. and D.R. Cox. 1995. A yeast assay for functional detection of mutations in the human cystathionine β -synthase gene. *Hum. Mol. Genet.* **4**: 1155-1161.
- Maschal, R.D., J. Koontz, and J. Sklar. 1995. Detection of mutations by cleavage of DNA heteroduplexes with a bacteriophage resolvase. *Nature Genet.* **9**: 177-183.
- Modrich, P. 1991. Mismatch repair. *Annu. Rev. Genet.* **25**: 229-248.
- Myers, R.M., L.S. Lerman, and T. Maniatis. 1985a. A general method for saturation mutagenesis of cloned DNA fragments. *Science* **229**: 242-247.
- Myers, R.M., S.G. Fischer, L.S. Lerman, and T. Maniatis.

FAHAM AND COX

- 1985b. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acid Res.* **13**: 3131-3145.
- Myers, R.M., N. Lumelsky, L.S. Lerman, and T. Maniatis. 1985c. Detection of single base substitutions in total genomic DNA. *Nature* **313**: 495-498.
- Myers, R.M., Z. Larin, and T. Maniatis. 1985d. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**: 1242-1246.
- Nelson, S.F., J.H. McCusker, M.A. Sander, Y. Kee, P. Modrich, and P.O. Brown. 1993. Genomic mismatch scanning: A new approach to genetic linkage mapping. *Nature Genet.* **4**: 11-18.
- Nielson, K., J. Braman, and K. Kretz. 1995. High fidelity long PCR amplifications with pfu DNA polymerase. *Strategies* **8**: 26-33.
- Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi, and T. Sekiya. 1989a. Detection of polymorphisms in human DNA by gel electrophoresis as single-strand conformational polymorphism. *Proc. Natl. Acad. Sci.* **86**: 2766-2770.
- Orita, M., Y. Suzuki, T. Sekiya, and K. Hayashi. 1989b. Rapid and sensitive detection of point mutation and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**: 874-879.
- Parker, B.O. and M.G. Marinus. 1992. Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **89**: 1730-1734.
- Perry, D.J. and R.W. Carrell. 1992. A rapid and simple approach to the detection of DNA mutations in thromboembolic diseases. *J. Clin. Pathol.* **45**: 158-160.
- Sheffield, V.C., D.R. Cox, L.S. Lerman, and R.M. Myers. 1989. Attachment of a 40-base-pair G + C rich sequence (G + C clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci.* **86**: 232-236.
- Short, J.M., J.M. Fernandez, J.A. Sorge, and W. Huse. 1988. λ ZAP: A bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**: 7583-7600.
- Wagner, R. and M. Meselson. 1976. Repair tracts in mismatched DNA heteroduplexes. *Proc. Natl. Acad. Sci.* **73**: 4135-4139.
- White, M.B., M. Carvalho, D. Derse, S.J. O'Brien, and M. Dean. 1992. Detecting single base pair substitutions as heteroduplex polymorphisms. *Genomics* **5**: 301-306.
- Wilson, B.D., M.M. Ollmann, L. Kang, M. Stoffel, G.I. Bell, and G.S. Barsh. 1995. Structure and function of ASP, the human homologue of the mouse agouti gene. *Hum. Mol. Genet.* **4**: 223-230.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequence of the M13mp18 and puc19 vectors. *Gene* **33**: 103-119.
- Youil, R., B.W. Kemper, and R.G.H. Cotton. 1995. Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl. Acad. Sci.* **92**: 87-91.

Received December 1, 1995; accepted December 5, 1995.

DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory

(screening for mutation/gel electrophoresis/helix-coil transition/ λ phage/restriction fragment)

S. G. FISCHER AND L. S. LERMAN

Center for Biological Macromolecules, Department of Biological Sciences, State University of New York, Albany, New York 12222

Communicated by Rollin D. Hotchkiss, December 23, 1982

ABSTRACT DNA fragments 536 base pairs long differing by single base-pair substitutions were clearly separated in denaturing gradient gel electrophoresis. Transversions as well as transitions were detected. The correspondence between the gradient gel measurements and the sequence-specific statistical mechanical theory of melting shows that mutations affecting final gradient penetration lie within the first cooperatively melting sequence. Fragments carrying substitutions in domains melting at a higher temperature reach final gel positions indistinguishable from wild type. The gradient data and the sites of substitution bracket the boundary between the first domain and its neighboring higher-melting domain within eight base pairs or fewer, in agreement with the calculated boundary. The correspondence between the gradient displacement of the mutants and the calculated change in helix stability permits substantial inference as to the type of substitution. Excision of the lowest melting domain allows recognition of mutants in the next ranking domain.

Detection and localization of single base substitutions within long DNA sequences may be impractical by complete sequence determination and improbable on the basis of restriction endonuclease vulnerability. We present here the results of a procedure by which DNA molecules that have minimal sequence differences are separated and by which some conclusions can be drawn as to the nature of the change. A number of samples can conveniently be examined in a single slab gel; each DNA species is focused into a sharp band at a gel position determined by its sequence and composition. The physical separation of fragments of altered sequence provided by the denaturing gel makes possible further analysis and manipulation.

Our system makes an unconventional use of electrophoresis. Where DNA molecules migrate into a gradient of ascending concentration of denaturant, they undergo an abrupt decrease in mobility at a characteristic depth, resulting in positions and patterns that change little if application of the field is continued. The retardation depth in the gradient is determined by the least stable part of the molecule and is relatively insensitive to other parts of the sequence or to the overall length (1).

To understand the basis of the sensitivity of the system to single base substitutions, we have undertaken a close comparison of the gel results with those of a sequence-specific statistical mechanical theory of the stability of the double helix. Experimental studies on the helix-disorder transition, melting, have not yet provided a detailed test of the theory, which predicts intricate and interesting patterns for the progression of equilibria from full helicity to separated strands as the temperature increases for molecules of different sequence. In our system, the molecule is exposed to a gradual denaturation-pro-

moting change in the medium, linearly equivalent (2) to a gradual increase in temperature. The strong decrease in mobility as the helix unravels provides the basis for both sequence-determined separation and examination of the helix-disorder transition.

MATERIALS AND METHODS

DNA Preparation. λ strains were obtained from D. Wulff as *Sam7* derivatives substituted in the y region. Strain identifications and sequence determinations are from Wulff *et al.* (3) and D. Wulff and M. Rosenberg (personal communications). DNA from those strains was prepared as described. DNA was also prepared from λ strain c1857 by lytic infection of *Escherichia coli* strain K-12 W3110 by standard methods. Plasmid pKM2 (K. McKenney and M. Rosenberg, personal communication), containing the wild-type λ sequence at position 38,989-40,291 inserted into the *Hind*III site of pBR322, was grown in *E. coli* HB101. Phage and plasmid DNAs were digested to completion with *Ava* I/*Bgl* II or *Alu* I/*Bgl* II under the conditions specified by the supplier (Bethesda Research Laboratories). Digestion was stopped by the addition of EDTA to a final concentration of 40 mM, glycerol to a final concentration of 10% (vol/vol), and a trace of bromphenol blue tracking dye.

Gel Electrophoresis. Nondenaturing polyacrylamide (65 mg/ml; acrylamide/bisacrylamide, 30:0.8) gels were run at 60°C in TAE buffer (40 mM Tris/20 mM NaOAc/1 mM EDTA, pH 8/HOAc), with 3-5 μ g of whole phage DNA or 0.5-1.0 μ g of plasmid DNA in each slot.

Denaturing Gradient Gels. The acrylamide concentration, 65 mg/ml, and the TAE buffer concentration were uniform throughout the gels. All gradients consisted of linearly increasing concentrations of urea/formamide at a constant ratio. Gels were poured from outgassed solutions containing ammonium persulfate at 0.1 mg/ml and 0.01% *N,N,N',N'*-tetramethylethylenediamine using a two-chamber gradient maker or the syringe gradient pump previously described (4, 5). The gels were submerged in a 5-liter aquarium that contained the anode electrolyte, TAE. The electrolyte was stirred and controlled at 60.0°C, and a field of 6 V/cm was applied.

RESULTS

Gradient Discrimination of Mutants. We have compared wild type and 16 strains of λ that have substitutions in the y region using fragments produced by cleavage with *Ava* I/*Bgl* II. The fragments contain 536 base pairs (bp) preceded by four unpaired bases from the *Bgl* II site. Numbering begins with the first paired base in the *Ava* I site. A map of the sites and substitutions for each of the mutants is shown in Fig. 1. Strain *cin-1 cnc-1* is a double mutation containing both *cin-1* and *cnc-1* substitutions.



FIG. 1. Sites of λ region mutants. According to the convention by which the *EcoRI* site between the fragments whose lengths are 7,421 bp and 5,806 bp is numbered 40,000, the first base of this fragment, indicated here as 1, would be numbered 39,125.

The substitutions *ctrl* and *ctr2* are present only in strains also carrying the substitution *cII3059*, and strain *cir5* contains both *cir5* and *cII3086* substitutions.

DNA of each strain was digested with *Ava I*/*Bgl II* and separated by length in a nondenaturing polyacrylamide gel. The corresponding 536-bp fragments migrate to the same depth in each lane. The narrow strip carrying the bands was sealed across the top of a gradient gel in which denaturant concentration increased from top to bottom in a uniform polyacrylamide matrix. A field of 150 V was applied for 14.5 hr at 60°C.

As shown in Fig. 2, which presents a set of similar gradient gels with the wild-type fragment and 16 mutants, substitution of T:A for C:G at position 129, strain *cy3019*, results in retardation 0.9 cm higher in the gel than for wild type. The inverse substitution at position 136, strain *cy3071*, delays retardation beyond that of wild type. The shifts are not fully defined by the composition of the substitution; *cy3019* is displaced almost twice as far as *cII3105*, although both represent replacements of G:C by A:T. The difference between *cy2001* and *cy3071* is a simple transversion, the interchange of cytosine and guanine across strands, but there is a more-than-2-mm difference in gel positions. The doubly substituted strain, *cin-1 cnc-1*, identical to wild type in gross composition in that it carries both A:T \rightarrow C:G and G:C \rightarrow T:A substitutions, is also shifted. A single T:A \rightarrow C:G substitution at position 142 in strain *cII3086* effects retardation below wild type while an additional A:T \rightarrow G:C substitution in strain *cir5* delays retardation to an even greater depth. All of these substitutions occur within the first 144 bp adjacent to the *Ava I* site. The four mutants with substitutions at base 153 and above have no detectable effect on the gradient position. Strain *cII3059* and its derivative double mutants are discussed below.

Effect of Single Base Substitution on the Mobility Shift. The basis for separation is shown by comparison of the electrophoretic mobilities of two fragments in gels in which both fragments migrate through a constant concentration of the denaturing solvent (4). The variation in mobility over a substantial range of denaturant concentration is displayed by a denaturing gradient perpendicular to the electric field. In this procedure, 20 μ g of wild-type λ *Sam7* DNA and 10 μ g of DNA from strain λ *Sam7cy2001*, containing an A:T \rightarrow C:G transition at position 136, were cleaved with *Ava I*/*Bgl II* and run into an agarose (15 mg/ml) slab gel from a single 13-cm-wide starting zone. An agarose strip containing the 536-bp band was cut from the ethidium-stained gel and sealed with hot agarose across the top of a polyacrylamide (65 mg/ml) gel containing a linear gradient from 1.4 M urea/8% (vol/vol) formamide on the left to 3.5 M urea/20% (vol/vol) formamide on the right. A field of 150 V perpendicular to the gradient was applied for 7 hr at 60°C (Fig. 3).

Fragments on the left migrate through a column of minimum denaturant concentration. From left to right, fragment mobility at first decreases gradually, then suffers a sharp reduction two-thirds of the way to the right at about 2.8 M urea/16% (vol/vol) formamide. At this point, the fragment from strain 2001 is distinguishable from wild type as a species that has half the staining intensity; its sharp mobility transition occurs at a very slightly higher denaturant concentration. The wild-type and mutant fragments appear to be superposed in both the high- and low-concentration regions where the mobility changes only slightly with increasing denaturant concentration. Fragment mobility decreases to less than one-fourth through the major transition zone.

The Calculated Melting Map. To examine the effect of a single base substitution on the calculated progression of the helix-

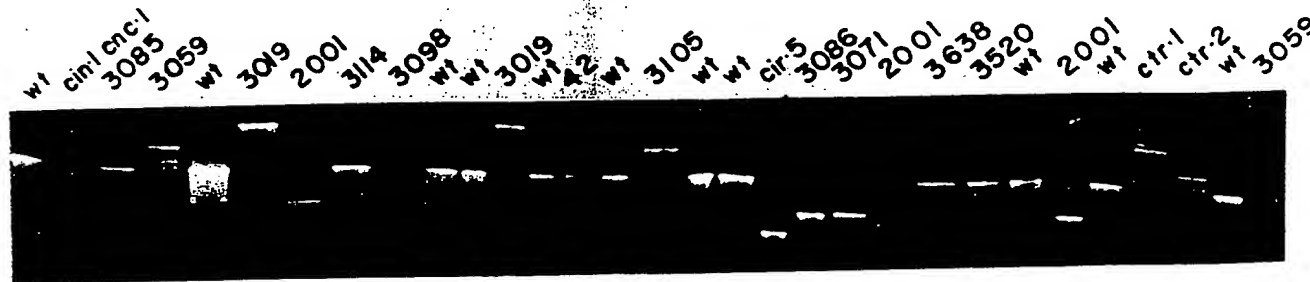


FIG. 2. Effect of mutation on fragment depth in denaturing gradient gels. Three to 5 μ g of λ *Sam7* DNA and derivatives substituted in the γ region (see Fig. 1) digested to completion with *Ava I*/*Bgl II* was electrophoretically separated in a nondenaturing polyacrylamide (65 mg/ml) gel. The ethidium-stained polyacrylamide strip containing the 536-bp γ -region fragment was sealed with hot agarose (10 mg/ml) in TAE buffer across the top of a polyacrylamide (65 mg/ml) gel containing a linear gradient of a urea/formamide mixture increasing, from 2.1 M urea/12% (vol/vol) formamide to 3.5 M urea/20% (vol/vol) formamide, in the same direction as the electric field. The figure is a composite of the central 2.5 cm of each of four gels, each containing one or more wild-type (wt) samples.

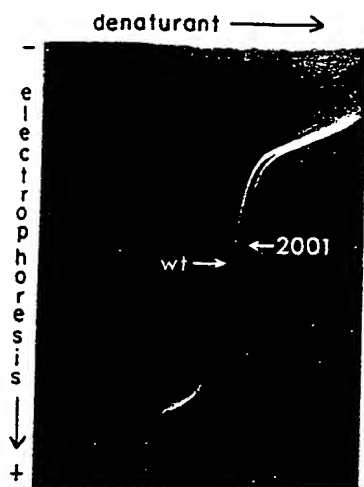


FIG. 3. Comparison of mobilities of wild-type and cy2001 fragments in a denaturing gradient perpendicular to the electric field. A mixture of wild-type and cy2001 fragments was electrophoresed in a gel in which the denaturant concentration was constant along the path of electrophoretic movement but increased linearly in the perpendicular direction. The sample was applied uniformly along the top from a strip of an agarose gel containing nominally the same total amount of both fragments at every point. The amount of mutant DNA was half that of wild-type DNA.

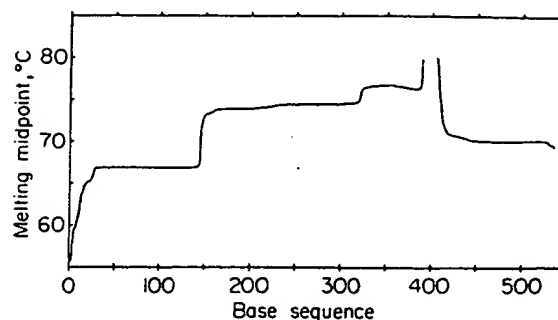


FIG. 4. Expected progression of melting of the wild-type λ fragment. The abscissa represents positions along the sequence between nearest-neighbor base pairs from the first pair following the *Ava* I scission position through the last pair at the *Bgl* II site. The distribution of each base pair between the helical and melted states was determined from the Fixman-Friere (6) algorithm at 0.1°C increments, and the temperature for 50:50 equilibrium was inferred by cubic interpolation. The function is not shown in the highest melting region between bases 389 and 405 where the properties depend on the probability of complete separation and bimolecular reassociation. $\sigma = 3.3 \times 10^{-5}$, $\delta(n) = \sigma n^{-2.0}$, approximated as a sum of nine exponentials; $\Delta S/R = 12.25$ mol. The Gotoh-Tagashira values for nearest-neighbor 5'-3' doublets in 19.5 mM Na⁺ are as follows: T-A, 36.73°C; T-T, 54.50°C; T-G, 86.44°C; A-T, 57.02°C; A-G, 58.42°C; A-C, 97.73°C; C-G, 72.55°C; C-C, 85.97°C; G-C, 136.12°C (8).

random chain equilibrium, we have used the Fixman and Friere (6) modification of the algorithm presented by Poland (7) for calculation of the equilibrium melting transition (T_m) probability as a function of sequence and temperature. We have replaced the 2-valued stability parameters for base pairs (usually given as T_{MAT} or T_{GCC}) by the set of 10 values for nearest-neighbor doublets suggested by Gotoh and Tagashira (8). The values are given in Fig. 4. A base-pair substitution changes the stability values in two adjacent positions, and a transversion without substitution results in a significantly different net value although the overall base composition remains constant. Since the Poland-Fixman-Friere calculation depends on at least two statistical parameters that are not precisely known (the cooperativity constant, σ , and a loop-closure exponent, α), we have compared results for several values and combinations of each. We have ascribed the variation in T_m of nearest-neighbor doublets to variation in ΔH , holding ΔS constant, but the converse assumption gives essentially the same results. The patterns of melting progression and the differences due to substitution show only insignificant differences so long as the parameters do not depart drastically from values previously published by others. All of the results shown here are based on $\sigma = 3.3 \times 10^{-5}$, the center of the range proposed by Amirikyan *et al.* (9), and $\alpha = 2.0$ (10).

The base sequence of the fragment enters the calculation of the melting progression as the sequence of nearest-neighbor stability values, represented as temperatures. The expected melting progression along the molecule calculated with standard parameters is shown in Fig. 4 as a melting map—the temperature at which each base will be at equilibrium with equal probability of helix or random chain configuration. At any temperature appreciably below the contour for that base pair, the pair can be regarded as helical and, at any temperature appreciably above the contour, the pair can be regarded as melted. As expected, a few base pairs at each end melt at lower temperatures; ends of the helix fray gradually prior to any large cooperative changes. The first domain to melt consists of the bases between 32 and 142. The uniformity of the ordinate value shows that all of the bases of this region melt as a block. A base at the

center of the block progresses from 10% to 90% probability of unpairing and unstacking between 66.3°C and 67.3°C. The calculated contour rises 0.3°C between bases 143 and 144, 2.8°C between bases 144 and 145, 1.5°C between bases 145 and 146, and 0.4°C between bases 146 and 147, giving the appearance of a steep wall. The last melting necessary for strand separation is omitted; it is concentration dependent and unnecessary for the present analysis. The calculation shows, in general, that melting of a DNA molecule can be expected to proceed stepwise as the temperature is raised; the effect of cooperativity is strong enough that fairly long blocks of contiguous helix melt within temperature intervals much narrower than the temperature differences between entire blocks. The separation into blocks, or domains, follows from the algorithm, and no assumptions or *a priori* estimates of domain boundaries are required. The loci of domain boundaries are not discernable by inspection of the sequence.

We note that the lowest melting domain contains the *cyL* region, the sequence thought to determine polymerase recognition, and the *cyR* region falls into a higher melting domain beginning at about base 147 (3).

The changes in the melting progression effected by mutations between bases 83 and 168 are presented in Fig. 5, together with a section of the standard map. The calculations are shown in Fig. 5B and C as the differences in temperature for the 50% point in the melting equilibrium between the wild-type fragment and each mutant at each base pair. The difference maps were obtained by subtracting the melting map of the complete sequence of the wild-type fragment from the corresponding complete map of each mutant. As shown in Fig. 5B for four representative substitutions and a double substitution, substitutions that effect substantial displacements in the gels distinctly alter the melting temperature in the entire first domain. Except for slight shifts in the domain boundaries in this set (a shift appears as a spike in the melting map), melting is essentially identical with that of wild type at all other parts of the sequence. The effect of the double substitution, *cin-1 cnc-1*, and the comparison between *cy2001* and *cy3071* are of particular interest because corresponding differences are seen both in the gel positions and in the calculation between domains of identical

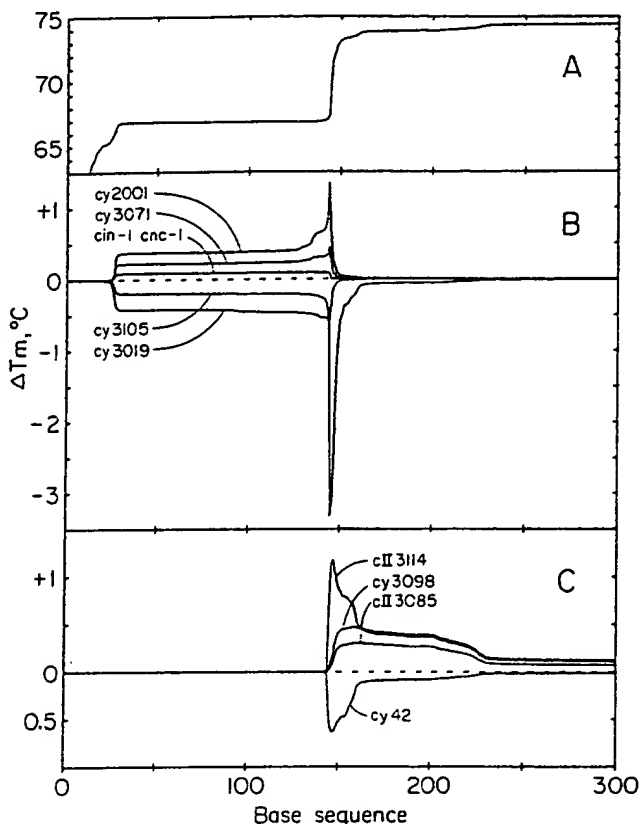


FIG. 5. Region of the first (lowest melting) domain and difference melting maps resulting from base substitutions. (A) The low-numbered region of the melting map of the wild-type fragment (Fig. 4) is shown with an expanded temperature scale. (B and C) T_m values for each base in the wild-type sequence have been subtracted from T_m values calculated for the mutant sequences.

base composition as a consequence of nearest-neighbor interactions.

The calculated melting difference maps for strains that reach retardation depths in the gel indistinguishable from the parental λ strain are shown in Fig. 5C. They have a negligible effect (less than 0.01°C) below base 145 in the first domain but depress the edge of the adjacent high-melting domain. The effects of these mutations tend to be significant from base 147 to about base 225. Substitutions at sites 424 in cII3638 and 433 in cII3520, show gradient positions indistinguishable from wild type and affect the melting map only near the high-numbered end.

Detection of Mutants Near the *Bgl* II End. The region within which substitutions in the 536-bp *Ava* I/*Bgl* II fragment are discernible in the gel corresponds to the lowest melting domain of the theoretical map. Excision of the low-numbered end of the sequence can be expected to promote the region of next higher T_m , extending from base 416 to the *Bgl* II end, to first melting, so that in the shortened fragment substitutions in this region should be recognizable. Gel positions of the truncated fragments of the wild type, cII3638, and cII3520, in which the same sequence was cleaved between bases 235 and 236 with *Alu* I, are shown in Fig. 6. Wild-type fragments from the recombinant plasmid pKM2 (lanes A and E) focused into a sharp band at the same depth as a band of the fragment from whole phage DNA (lane B). The mutant fragments, indistinguishable from wild-type fragments in the gradient in the original 536-bp molecule (Fig. 2) focus at greater depths, in good agreement with the calculated effect of the substitutions based on melting and mobility theory. Note that all of the fragments derived from whole phage DNA,

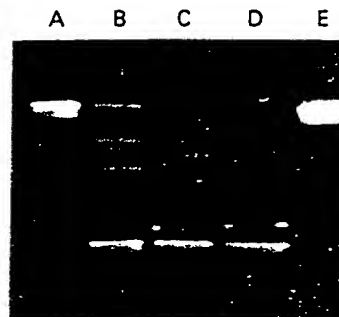


FIG. 6. Gel positions of truncated fragments substituted at the high-numbered end. The 301-bp *Alu* I/*Bgl* II fragments from pKM2 (lanes A and E), Sam7 (lane B), cII3638 (lane C), and cII3520 (lane D) were analyzed on a 42–60% denaturant/polyacrylamide (150 mg/ml) gel. The electric field was applied for 14 hr.

mutants and wild type, appear as triplet bands in which the substitutions uniformly shift all three members. A singlet lower in the gel, unaffected by the substitutions, provides a reference position. Tripletting of the 301-bp *Alu* I/*Bgl* II fragment was also obtained from wild-type phage grown lytically, while the plasmid-derived singlet was unaffected by mixed digestion with whole phage DNA. These results are consistent with the supposition that base modification during phage growth may be a source of the extra bands.

DISCUSSION

Because the gradient interval between any pair of fragments differing by a single base substitution in the determinant domain is larger than the width of the bands, samples can be recovered from the gels enriched for either component. The bands are narrower than those in simple length separations by constant-velocity electrophoresis because of the focusing due to the reduction in mobility as the determinant domain melts.

The correspondence between these results and properties calculated from the sequence by the Poland-Fixman-Friere algorithm provides more detailed support for this melting theory than has been available from hyperchromicity profiles. Because the six substitutions that alter gradient depth lie below position 145 and the six that do not alter gradient depth lie above position 152, it appears that the sequence above position 152 does not participate in determining the retardation depth. Following the explanation we have offered for identification of the decrease in mobility with partial melting (1, 11), we infer that the low-numbered region ending between bases 144 and 152 melts at a substantially lower temperature and independently of melting above base 152. That difference and the resulting decoupling emerges from the theoretical calculation as a distinct domain boundary within the limits specified by these mutants. The loss of mobility from melting of the first domain prevents the fragment from reaching the gradient depth necessary for further melting within the duration of the run.

Variation of the loop-entropy and cooperativity parameters can result in melting at a slightly lower temperature in the region between bases 83 and 131 than in the region near base 61, but there is no noticeable change in the maps shown in Fig. 5B nor in the boundary position.

There is a close proportionality between the calculated alteration, ΔT_m , of the plateau T_m of the first domain and the change in gradient depth reached by the fragments (Fig. 7). For sequences of different domain lengths, a simple comparison according to domain T_m is not appropriate and, hence, the sequence carrying the double substitution *cir5*-cII3086, which shifts the domain boundary about six bases to the *Ava* I end, is omit-

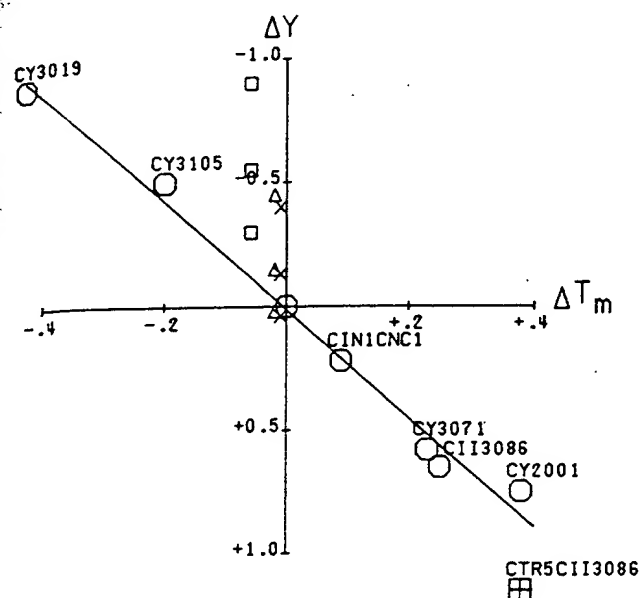


FIG. 7. Gel displacement and alteration in the T_m of the lowest melting domain. ΔT_m is the temperature difference between each mutant and wild type at the center of the first domain. ΔY is the increment in depth in the gradient referred to wild. The point at the origin represents wild type and seven strains carrying substitutions above base 152. The mutants that give triplet zones are as follows: Δ , cII3059; \square , cII3059 *ctr1*; \times , cII3059 *ctr2*; they have not been included in the calculation of the regression line.

ted from the regression line. The level of consistency between the calculated T_m of the first domain and the observed differences in gel depth support the relative assignments of nearest-neighbor stability values proposed by Gotoh and Tagashira (8) on the basis of less direct evidence. Our use of these values is somewhat arbitrary, in that they were proposed as applicable to conditions differing from those in the gels—0.02 M aqueous Na^+ , rather than 0.02 M Tris^+ and urea/formamide. However, the doublet stabilities always enter the calculation as the sum of nearest-neighbor values from both sides of each base pair, and the present results remain compatible with other doublet-stability assignments. Gel measurements with a large set of mutants may constitute an appropriate means to infer nearest-neighbor stability assignments.

Calculation of the melting map using only 2 stability values, one for G-C pairs and another for A-T, without consideration of neighbors results in a similar melting map. However, the 2-valued calculation cannot account for the gel displacement of the double mutant *cin-1 cnc-1*, which has a base composition identical to that of the wild type, nor for the difference due to transversion found in the comparison between *cy2001* and *cy3071*.

While most bands are accompanied to some extent by weak satellites slightly deeper in the gradient, the relative intensities of the satellites from cII3059 and its two derivatives are conspicuously greater. This property is retained through repeated plaque purification. Since the multiplicity also appears distinctly where the cII3059 fragment moves through a constant denaturant concentration, giving a pattern similar to that shown

in Fig. 3, the multiplicity is not generated by details of the gradient. The DNA in each band of the triplet behaves as a stable single component after isolation and migration into a new gradient. The calculated effect of the nominal base substitution in cII3059 appears in the difference melting map almost entirely as a perturbation beyond the boundary of the first domain. We have been unable to arrive at a significantly larger change in the first domain despite substantial *ad hoc* adjustment of many of the nearest-neighbor stability values by 5–20°, either singly or in combination, or by changes in other parameters. If the most displaced (highest) bands of the cII3059 mutants are related to the principal wild-type band, this effect could be construed as an influence on mobility originating from mutation about 6 bp pairs beyond the calculated boundary. If the least shifted (lowest) component is taken to indicate the effect of substitution, the displacements are compatible with the calculated domain boundary, as shown by the cluster of points near the origin in Fig. 7. The interval of 2 bp between cII3105 and cII3059 corresponds to the steepest part of the calculated boundary.

These results suggest that all substitutions from point mutations can be detected in denaturing gradient gels if they occur in the lowest melting domain of the molecule. Sensitivity will depend on the net stability change, which depends, in turn, on the specific context of the substitution, and on the length of the domain over which the effect is averaged. Perturbations at least as large as those from substitutions can be expected from single base pair insertions and deletions and from some base modifications. In the present example, the sensitive region initially constitutes 20% of the 536-bp fragment; it can be shifted to a different 20% by cleavage, which transfers highest melting priority to a different domain. Since the separations do not critically depend on the length of the molecules, random fragmentation may be as useful as restriction cleavage. By attachment of a higher melting section to small restriction fragments, nearly any sequence of appropriate length can be made to constitute the lowest melting domain in the resulting composite molecule, and nearly all substitutions may be made discernible.

We wish to thank Daniel Wulff and Michael Mahoney for strains from their collection of λ mutants. We are grateful to Karen Silverstein, Michele Soberman, Christine Cullen, Arthur Eisenberg, and Nelly Brown for assistance and to Marshall Fixman, Nadrian Seeman, and Richard Blake for helpful discussions. This work was supported by Grant GM-24030-06 from the National Institutes of Health.

1. Fischer, S. G. & Lerman, L. S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 4420–4424.
2. Lerman, L. S., Fischer, S. G., Bregman, D. B. & Silverstein, K. J. (1981) in *Biomolecular Stereodynamics*, ed. Sarma, R. H. (Adenine, New York), pp. 459–470.
3. Wulff, D. L., Behr, M., Izumi, S., Beck, J., Shimatke, H., Brady, C., Court, D. & Rosenberg, M. (1980) *J. Mol. Biol.* 138, 209–230.
4. Fischer, S. G. & Lerman, L. S. (1979) *Cell* 16, 191–200.
5. Fischer, S. G. & Lerman, L. S. (1979) *Methods Enzymol.* 68, 183–191.
6. Fixman, M. & Friere, J. J. (1977) *Biopolymers* 16, 2693–2704.
7. Poland, D. (1974) *Biopolymers* 13, 1859–1871.
8. Gotoh, O. & Tagashira, Y. (1981) *Biopolymers* 20, 1033–1043.
9. Amirikyan, B. R., Volodgodshii, A. V. & Lyubchenko, Y. L. (1981) *Nucleic Acids Res.* 9, 5469–5482.
10. Blake, R. D. & Fresco, J. R. (1973) *Biopolymers* 12, 775–789.
11. Lerman, L. S. & Frisch, H. L. (1982) *Biopolymers* 21, 995–997.

Detection of single DNA base differences by competitive oligonucleotide priming

Richard A. Gibbs¹*, Phi-Nga Nguyen² and C. Thomas Caskey^{1,2}¹Institute for Molecular Genetics and ²Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA

Received January 4, 1989; Accepted February 23, 1989

ABSTRACT

Synthetic DNA oligonucleotides can serve as efficient primers for DNA synthesis even when there is a single base mismatch between the primers and the corresponding DNA template. However, when the primer-template annealing is carried out with a mixture of primers and at low stringency the binding of a perfectly matched primer is strongly favored relative to a primer differing by a single base. This primer competition is observed over a range of oligonucleotide sizes from twelve to sixteen bases and with a variety of base mismatches. When coupled with the polymerase chain reaction, for the amplification of specific DNA sequences, competitive oligonucleotide priming provides a simple general strategy for the detection of single DNA base differences.

INTRODUCTION

Techniques enabling the rapid detection of single DNA base changes are important tools for genetic analysis (1). When the precise DNA base change in a mutation is known allele specific oligonucleotides (ASO's) can identify the unique sequences by differential hybridization under stringent conditions (2). Typically, 18–20 base oligonucleotides are constructed with perfect complementarity to either a normal (wild-type) or mutant sequence. The DNA for analysis is tethered to a solid support and hybridized separately to the radioactive probes. The homology between each ASO and the test sequence is then revealed by sequential washings of the hybrids at high stringency so that the mismatched probe is washed free while the perfect match remains bound.

In order to simplify the detection of single DNA base changes we have used an alternative strategy employing mixtures of synthetic DNA oligonucleotides as primers for DNA synthesis. An example of the basic principle is outlined in Fig. 1. Two synthetic oligonucleotide primers are mixed in a single annealing reaction with a DNA template. Each of the primers is capable of priming DNA synthesis at the same site. However, when one primer is perfectly complementary to the DNA template it can bind in preference to a primer that differs by a single base. The use of a third oligonucleotide primer (common primer) allows the identification of the successfully competing primer by the polymerase chain reaction (PCR) (3–6). When the primer-annealing reactions are carried out at low stringency and with an excess primer to template ratio, the perfectly matched primer can be favored with a level of discrimination greater than 100:1. The perfectly matched primer will also compete successfully when it is at low abundance (*i.e.*, in the presence of up to a 100-fold excess of a mismatched primer) or when the correct match is only one of a four member mixed oligonucleotide family. We also show here that successful competitive oligonucleotide priming (COP) is dependent on the length of the primers and that the COP

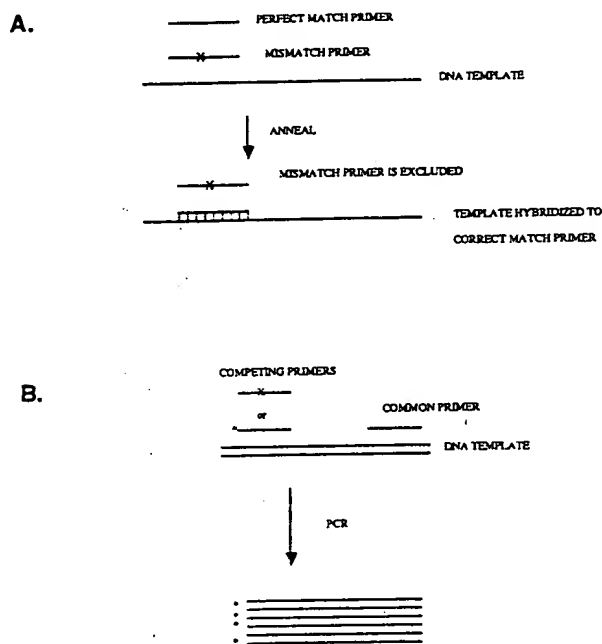


Figure 1. Overview of the general strategy for detection of single DNA base differences by competitive oligonucleotide priming (COP). **A.** DNA template is mixed with two oligonucleotides that differ by a single DNA base. If one oligonucleotide is a perfect match to the DNA template it will bind in preference to a mismatched oligomer. **B.** The correct match primer may be identified by differentially labeling the two oligonucleotides and including a third 'common' oligonucleotide primer. The common primer and the 'successful' COP primer are incorporated into a DNA fragment generated by PCR. Identification of the incorporated COP primer infers the template sequence.

system can be coupled to the PCR to detect single DNA base changes in mammalian genomic DNA.

MATERIALS AND METHODS

Oligonucleotide primers (Table 1) were synthesized on an Applied Biosystems 380B oligonucleotide synthesizer using β -cyanoethyl phosphoramidite chemistry. Mixed oligonucleotides were synthesized using the 380B mixed (competitive) coupling functions. The relative efficiency of addition of individual bases during mixed synthesis has been the subject of previous reports by the manufacturer (7). Following synthesis, oligonucleotides were deprotected in ammonium hydroxide for 6–12 hours at 55°C, dried, dissolved in formamide and purified by denaturing polyacrylamide gel electrophoresis. The oligonucleotides were electroeluted from gel slices and finally desalted over an NENsorb column (Dupont).

The PCR for the amplification of the competitive priming events was carried out either with the large fragment from *E. coli* DNA polymerase I (Klenow) (United States Biochemical Corporation; USB) or with the heat stable DNA polymerase from *Thermus aquaticus* (Taq) (Perkin Elmer/Cetus). Klenow reactions were in a final volume of 100

TABLE 1. OLIGONUCLEOTIDE PRIMERS

Number	Sequence ^a (5' to 3')	Length	Template
1	CCCAGTCACGACGTT	15	M13 'Common'
85	AGCTCGGTACCC	12	M13 polylinker
86	AGCTCGG(TA)AC(CG)C	12	"
98	CGAGCTCGG(TA)AC(CG)C	14	"
99	TTCGAGCTCGG(TA)AC(CG)C	16	"
100	AATTCGAGCTCGG(TA)AC(CG)C	18	"
89	AATTCGAGCTCGGTACCCGG	20	"
90	AATTC(GC)AGCTCGG(TA)AC(CG)CGG	20	"
92	CAAGTGAATGTC	12	OTC mutation (-)
93	CAAGTGAATGTC	12	OTC mutation (sp)
94	CTGTCCACAGAAACAGGC	18	OTC 'Common'
246	GGCGATGTCAATAGGACTCCAGATG	25	HPRT genomic
352	CCACGAAGTGTGGATATAAGC	22	HPRT genomic
383	TAATGACACAAACATG	16	HPRT Mutation (+)
384	TAATGACATAAACATG	16	HPRT Mutation (-)

^aParentheses denote mixtures of bases at a single position

μ l containing 30 mM Tris-acetate, pH 7.9, 60 mM sodium-acetate, 10 mM magnesium-acetate, 10 mM dithiothreitol, 1.5 mM each of dATP, dCTP, dGTP, dTTP, 4 μ M of each primer (or primer family) and 0.5 to 1.0 μ g of DNA template (3). To initiate the Klenow catalyzed PCR reactions, DNA was denatured in 0.4 N NaOH for 5 minutes at 25°C, neutralized with 1/10 volume of 2M ammonium-acetate, pH 4.5, and precipitated with 2.5 volumes of ethanol. The pellet was resuspended in the PCR mix and annealed at 28°C for 3 min., before the addition of 5 units of enzyme. The PCR proceeded with 2 min. polymerization at 28°C, 2 min., denaturation at 105°C (in a heat block containing glycerol) and 30 sec. annealing at 28°C before fresh enzyme was added. The Taq PCR followed the procedure of Kogan *et al.* (5) except that the concentration of each primer was 1.0 μ M. Temperature cycling of 37–55°C, 30 sec.; 65°C, 3 min; 92°C, 1 min. was controlled by an automated thermocycler (Perkin Elmer/Cetus).

Oligonucleotide primers were labeled at the 5' terminus with T4 polynucleotide kinase (USB) to a final specific activity of 25 Ci/mMol. PCR products were analyzed on either a 4% NuSieve agarose (Marine Colloids) or 12% polyacrylamide gels and dried for autoradiography. Plasmid DNA was prepared by two rounds of cesium chloride/ethidium

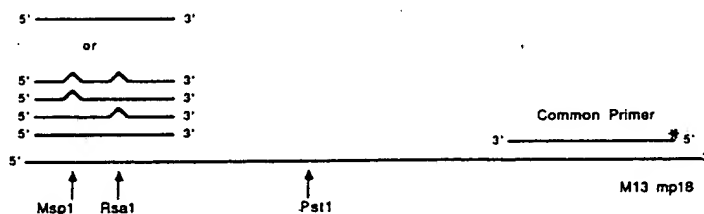
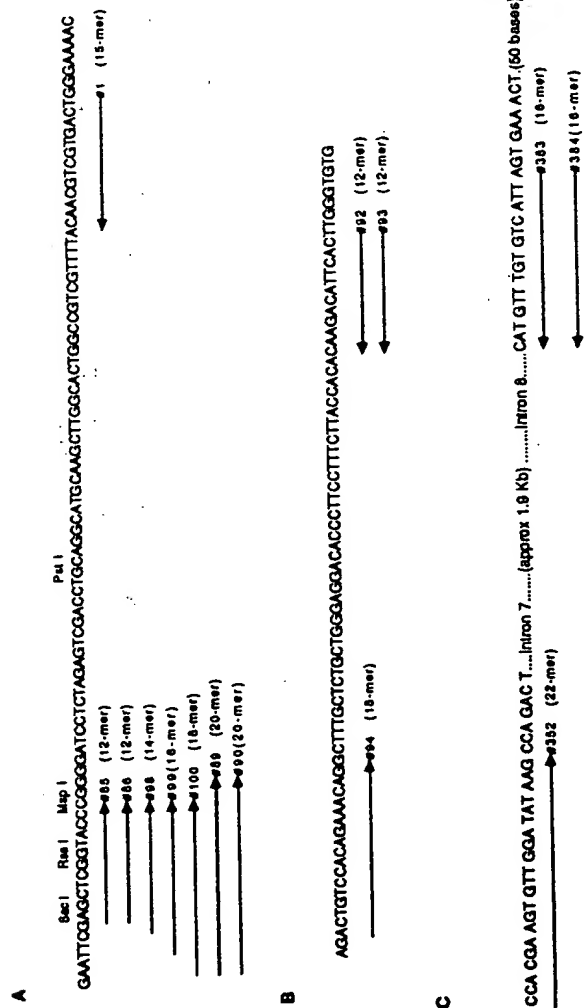


Figure 2. Strategy for analysis of competition between closely related oligonucleotide primers for an M13 (mp18) DNA template. The M13 multiple cloning site was PCR amplified with a radiolabelled (*) 'common' primer and either a perfectly matched opposing primer or a mixture of opposing primers that included single base mismatches to the DNA template. Incorporation of mismatched primers would lead to loss of restriction endonuclease recognition sites.



bromide density gradient centrifugation and human genomic DNA was prepared from transformed human lymphoblasts using an Applied Biosystems 340A DNA extractor.

RESULTS

Competition Between Oligonucleotide Primers of Different Length

Single stranded DNA (ssDNA) from the filamentous phage M13 (mp18) (8) was employed to first demonstrate the efficiency of the COP method. The scheme is illustrated in Fig. 2. The general strategy was to amplify a region of the mp18 multiple cloning site (polylinker) with primers that would either faithfully copy the restriction endonuclease recognition sequences or destroy the sites because of incorporation of oligonucleotide primers that included mismatches to the mp18 template. The DNA sequence of the template in the region bound by the competitive and common primers is shown in Fig. 3a.

The 'common' primer (#1, Table 1, Fig. 3) was a 15 base oligonucleotide that is often used as a universal DNA sequencing primer. Primers #85, 86, 89, 90, 98, 99 and 100 each overlap the opposite end of the mp18 polylinker where the *SacI*, *RsaI* and *MspI* restriction endonuclease recognition sites are located. Each of the primers were constructed so that they would be complementary to the products of extension of primer #1 through the mp18 polylinker. Primers #85 (12-mer) and #89 (20-mer) were completely homologous to mp18, and therefore could provide perfect copies of the ssDNA template. In contrast, primers #86, 98, 99, 100 and 90 were constructed as mixtures. Each contained two or more positions at which two nucleotides were added during synthesis. For example, primer #86, contained a mixture of A and T at position eight and C and G at position eleven. Thus, primer #86 had a complexity of four members, one with complete homology to the corresponding region of mp18, two with base mismatches (A:A, G:G) that altered single restriction endonuclease recognition sequences (*RsaI* or *MspI*) and one that altered both the enzyme sites.

Primer #1 was radiolabeled at the 5' terminus and employed in separate PCR reactions with each of the other primers and mp18 ssDNA template. The reaction products were either directly analyzed by gel electrophoresis and autoradiography, or first digested with a restriction endonuclease that identified a site within the amplified region. As expected the primer pair #1/85 generated an 85 bp fragment that was able to be digested to completion by *PstI* yielding a 48 bp radiolabeled fragment and by *RsaI* or *MspI* generating 77 or 74 bp products, respectively (Fig. 4a). Thus, the perfectly matched primer #85 was incorporated and faithfully reproduced the restriction endonuclease recognition sites from within its sequence. Somewhat surprisingly an identical result was obtained using the primer mixture #1/86. As only 25% of the primer #86 family is expected to be perfectly homologous to the mp18 template the presence of *RsaI* and *MspI* recognition sites within the amplified product indicated preferential incorporation of the perfectly matched primer relative to the family members with single or double DNA base mismatches. The apparent discrimination afforded by the competition for the correct match (*i.e.*, the relative incorporation of the perfect match vs a mismatch) is greater than 100:1 and even a prolonged

Figure 3. DNA sequence of the DNA templates used to analyze primer competition. Arrows showing the oligonucleotide primers point 5' to 3'. The sequence of the individual oligonucleotides are shown in Table One. A. M13 mp18 polylinker region. B. Murine ornithine transcarbamylase cDNA (9) C. Human hypoxanthine phosphoribosyltransferase (HPRT) exon sequences. The primers for the amplification of the human genomic DNA were each complementary to exon sequences.

Oligonucleotides #98, 99 and 100 were constructed to represent families of 14-, 16 and 18-mers with a complexity of four members each. Digestion of the products of PCR amplification of each of these in conjunction with primer #1 and the mp18 template revealed that the correctly matched oligonucleotides were predominantly incorporated although a small amount of material that was resistant to enzyme digestion was generated from the 18-mer family (Fig. 4c). We conclude that effective competition by a correct match primer may occur within a family of 12-mers, that the discrimination is greatly reduced within families of 20-mers and predict that the precise relationship between length and discrimination will vary depending on individual base mismatches and their sequence context.

Competitive Oligonucleotide Priming for Mutation Detection
The COP system was adapted for the identification of a known DNA single base change that leads to ornithine transcarbamylase (OTC) deficiency in the sparse fur (*spf*) mouse (9). Oligonucleotides were synthesized to overlap the C to A transversion (Table 1. Fig.



B.(20-mers)

Perfect Match (#1/89)					Mixed Primers (#1/90)				
not amplified	uncut	Pst1	Sac1	Rsa1	not amplified	uncut	Pst1	Sac1	Rsa1
				Msp1					Msp1

C.(14, 16, 18-mers)

Mixed Primers (#1/98)					Mixed Primers (#1/99)					Mixed Primers (#1/100)				
not amplified					not amplified					not amplified				
uncut					uncut					uncut				
Pst1					Pst1					Pst1				
Rsa1					Rsa1					Rsa1				
Msp1					Msp1					Msp1				

Figure 4. Competition between closely related primers for the M13mp18 single stranded DNA template. A complete explanation is given in the text. The results of restriction endonuclease digestion of amplified M13 DNA using A. 12 base, B. 20 base and C. 14, 16 and 18 base oligonucleotides as COP primers are shown. The incorporation of mismatched oligonucleotides leading to loss of a restriction endonuclease recognition sites is observed when the competing oligonucleotides are 20-mers (B., last three lanes).

3b). Primers #92 and 93 are 12-mers that match the wild-type and *spf* sequences, respectively and #94 is a common 18-base primer complementary to the OTC cDNA in an opposite orientation to primers #92 and 93. Ten cycles of PCR were performed with cloned wild-type OTC cDNA or the *spf* OTC cDNA as template using primers #94 and an equimolar mixture of #92 and 93. A trace of radiolabeled primer #92 or 93 was used to monitor the competition between the two oligonucleotides.

Fig. 5a shows an ethidium bromide stained 4% NuSieve agarose gel used to analyze PCR products generated from primers #94/92/93. A fragment of the predicted size is

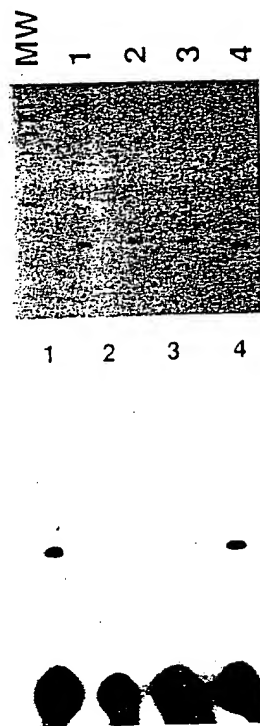


Figure 5. Identification of a C to A transversion in cloned murine ornithine transcarbamylase (OTC) cDNA. Four COP reactions containing competing primers complementary to the normal and mutant OTC sequences (# 92/93) and a common primer (# 94) (Table One) were analyzed by agarose gel electrophoresis. **Upper Panel:** Ethidium bromide-stained agarose gel electrophoresis of COP products (negative image). 1, normal template, normal match primer labeled, 2, normal template, mutant match primer labeled, 3, mutant template, normal match primer labeled, 4, mutant template, mutant match primer labeled. **Lower Panel:** Autoradiograph of the dried agarose gel, showing that the radiolabeled oligonucleotide primers were only incorporated in the presence of perfectly matched DNA templates.

generated from both the wild-type and the mutant templates. When this gel was dried and exposed to X-ray film only the fragments that were generated in reactions containing a radiolabeled primer that was a perfect match to the template had incorporated radioactivity (Fig. 5b). A prolonged exposure revealed faint bands from mismatch incorporation but indicated that the level of discrimination was greater than 100 to 1.

Preference for the 'correct' oligonucleotide primer

The strong preference of a DNA template for a 'correct' primer was further demonstrated by competing a perfectly-matched oligonucleotide primer with an excess of a mismatch oligonucleotide (Fig. 6). The preference of the cloned wild-type OTC template for the perfect match primer # 92 above # 93 when the two oligonucleotides were present in equimolar ratios (Lanes 1 and 2) was reduced only slightly when the radiolabeled mismatch (# 93) was present in a 100-fold molar excess (Lane 3). At a 1000:1 ratio, where the molar concentration of the correct primer was approximately that of the DNA template (Lane 4) the mismatch was incorporated with a still lower efficiency than when the mismatch

1 2 3 4 5 6

Figure 6. Successful primer competition in the presence of an excess of mismatched primer. The experiment is similar to that illustrated in Fig. 5 and a complete description is given in the text. Lanes 1 and 2 show incorporation and exclusion of radiolabeled perfect match and mismatched primers, respectively. Lane 3 indicates the level of incorporation of a radiolabeled mismatched primer present in a 100-fold excess above a perfectly matched, unlabeled primer (4.0 μ M vs 40 nM); Lane 4, mismatch present at 1000-fold excess (4.0 μ M vs 4.0 nM); Lane 5, mismatch alone (4.0 μ M). Lane 6 shows the exclusion of a radiolabeled mismatched primer (4.0 μ M) that was annealed to the DNA template for 3 min before addition of an equimolar amount of the correct match primer and initiation of the reaction.

primer was present alone (Lane 5). In a further reconstruction experiment the template was annealed to the radiolabeled mismatched oligonucleotide for the usual 3 min. before an equimolar amount of the correct primer was added. DNA polymerase was added after 1 minute more annealing and the PCR was carried out as before. Surprisingly, the correct match primer was predominately incorporated (Lane 6) reflecting the ability of the perfectly matched primer to displace any mismatched primer that might have been bound.

Detection of Single DNA Base Differences in Genomic DNA Using Taq DNA Polymerase
To test whether the COP mutation detection system could be adapted to conditions that enable the use of the heat stable Taq DNA polymerase (5,6), oligonucleotides complementary to normal or mutant human hypoxanthine phosphoribosyltransferase (HPRT) sequences were constructed (Table 1, ref 9). The oligonucleotides (16-mers) differed by a single base (C vs T) at the eighth position and had previously been employed as ASO hybridization probes to identify the corresponding normal and mutant alleles in a family study of HPRT deficiency (10). To enable COP analysis of the G to A transition an approximately 1950 base region of the HPRT gene containing the known mutation site was first PCR-amplified from genomic human DNA samples using primers #246/352. This fragment appeared homogeneous when analyzed by agarose gel electrophoresis. Five percent of the initial reaction products were then taken to initiate a further 10 rounds of PCR, with each of the allele-specific COP oligonucleotides (#383/384) and a common (#352) primer present. Four COP reactions were performed with either the correct match or mismatched primers radiolabeled at the 5' terminus and with either the normal or mutant 'preamplified' alleles as DNA templates. Analysis of each of the COP reactions by agarose gel electrophoresis and ethidium bromide staining revealed predominant products of the expected size. When the gel was dried and exposed to X-ray film it was found that the only fragments that were radiolabeled were the expected sized products from reactions where the radiolabeled primers perfectly matched the DNA templates. Thus the normal and mutant alleles were each correctly identified (Fig. 7).

DISCUSSION

We have demonstrated competition between closely related synthetic oligonucleotide primers

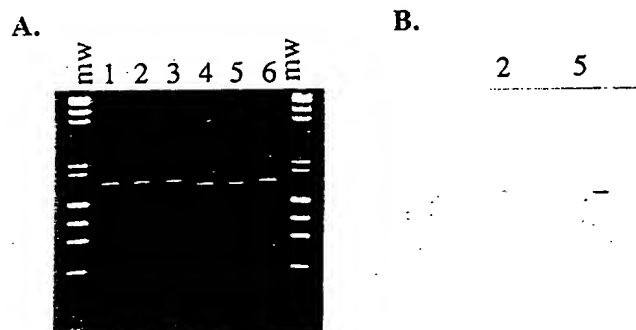


Figure 7. Identification of a single DNA base change in human DNA by COP, using Taq DNA polymerase. Four COP reactions containing competing primers complementary to normal and mutant HPRT gene sequences (#383/384) and a common primer (#352) (Table One) were performed using 'preamplified' human genomic DNAs as templates. **Left:** Ethidium bromide-stained agarose gel electrophoresis of PCR and COP products. Lanes 1 to 6 show the products of PCRs used to 'preamplify' the regions surrounding the normal and mutant alleles, respectively. Lanes 2-5 show COP products. 2, normal template, normal match primer labeled. 3, normal template, mutant match primer labeled. 4, mutant template, normal match primer labeled. 5, mutant template, mutant match primer labeled. **Right:** Autoradiograph of the dried agarose gel, showing the preferential incorporation of perfectly matched radiolabeled oligonucleotide primers.

for a single DNA template. When coupled to PCR, COP provides a simple method for the identification of single DNA base differences and thus represents an alternative to ASO probing for the analysis of mutations for which the precise DNA sequences can be predicted. In contrast to ASO probing the COP procedure does not require the use of solid filter supports and is technically more simple to perform. The COP strategy may therefore be favored for the routine analysis of known single DNA base substitutions or polymorphisms.

Successful COP has been shown to occur when competing oligonucleotides differ by single T-G, T-A, C-T or C-G base changes that generate A-A, G-G, G-A, T-C, C-A and T-G mismatches between the oligomers and their corresponding DNA templates. This represents six of the twelve possible base mismatches that may perturb normal Watson-Crick DNA base pairing. If reciprocal mismatches are equivalent (*e.g.*, A-T vs T-A) then only T-T and C-C mispairings are not described in this study. In other experiments C-C and G-G mismatches have been identified by COP (J. S. Chamberlain, *personal communication*) but it is likely that the general efficiency of the primer competition will be determined by both the mismatched base that is involved and its surrounding sequence context. Therefore many mutations in different sequences may need to be examined before a base mispairing that cannot be identified by COP could be found. The M13mp18 DNA template amplification/restriction strategy described here, coupled with mixed oligonucleotide synthesis represents a convenient method for further study of these relationships.

The competing oligonucleotide primers described above are short (12-16 nucleotides). Although this length exceeds that necessary for efficient priming of DNA synthesis it is less than the usual length employed for ASO probing. The efficiency of the competition can be reduced when the oligonucleotide primers are 20-mers but COP is still effective with 16-mers. As a general rule we are continuing to construct 16 base oligomers for

mutation identification as they can function both as ASO probes and competing primers, to further test the method.

An important feature of this study was the adaptation of Taq DNA polymerase to the COP reactions. If COP is to be a favored method of mutation detection then the procedure should be more simple to perform than alternatives that offer the same specificity of DNA sequence discrimination. The substitution of Taq in place of Klenow and the accompanying development of automated thermocyclers has enabled the widespread acceptance of PCR technology. The demonstration that a single DNA base difference can be identified by COP using Taq allows a similar protocol for point mutation detection to that used for routine PCR DNA amplification. The region containing the mutant sequence is first amplified and a small aliquot of the PCR product taken to initiate the COP reaction. The radiolabeled COP products are then analyzed by gel electrophoresis and autoradiography to identify the individual alleles.

A possibly important determinant of the relative hybridization efficiency of competing oligomers may be the position of the individual base mismatches within the oligonucleotides. In addition, the occurrence of a mismatch at the 3' terminus of an oligonucleotide may inhibit primer extension, provided the DNA polymerase used lacks a 3' to 5' exonuclease proofreading activity. Although this schema offers a strategy for mutation detection that is similar to COP in the manipulations and reagents that are required the underlying mechanism would be fundamentally different. In that case an oligonucleotide primer would not be required to bind specifically to the correctly matched allele, which contrasts to the central feature of the COP mechanism.

A mutation detection technique requiring a 3' base mismatched oligonucleotide and DNA ligase has been recently described (11). The method allows mismatch detection by failure of head-to-tail ligation of two synthetic oligonucleotides at the site of a mutant DNA base. Excellent discrimination between each of the 12 possible base mismatches and the corresponding perfect matches has been reported. The ligation method has a similar advantage to COP in that solid filter supports are not necessarily required. Unlike the primer competition reactions the ligation conditions are not so easily compatible to PCR buffers and therefore more extensive sample manipulation may be required for the analysis of PCR amplified DNA sequences.

There are many technical refinements that could be adapted to improve the COP method. More than two oligonucleotides can simultaneously compete for the same DNA priming site and multiple fluorescently labeled oligonucleotides (12) could be used to simultaneously test regions with a high degree of genetic heterogeneity. The maximum number of oligonucleotide species that could be employed in a single reaction has not yet been established but the observation that competition can occur when a mismatched primer is present at 100-fold higher abundance than the correct match suggests that even very highly polymorphic loci may be amenable to single analyses. Addition of a biotin residue to the 5' terminus of a common primer (13) could further facilitate the method by allowing rescue of a PCR-amplified fragment that has incorporated the successfully competing, differentially labelled primer by an avidin bound support. In this case the final analysis of the COP products would not require gel electrophoresis and could be monitored by measurement of the radioactive or fluorescent incorporation of the support matrix. Such refinements may eventually lead to the complete automation of DNA base difference detection for genetic disease diagnosis and the analysis of other DNA sequence polymorphisms in complex genomes.

ACKNOWLEDGEMENTS

We thank John Belmont, Jeff Chamberlain and Grant MacGregor for valuable discussions, Donna Muzny and Andrew Civitello for technical assistance and Elsa Perez for assistance in preparing the manuscript. R.A.G. is a recipient of the Muscular Dystrophy Association's Robert G. Sampson Distinguished Research Fellowship and C.T.C. is an investigator of the Howard Hughes Medical Institute. Supported by DHS grant # DK31428 and Welch Foundation grant # Q-533.

*To whom correspondence should be addressed

REFERENCES

1. Caskey, C.T. (1987) *Science*, **236**, 1223-1229.
2. Kidd, V.J., Wallace, R.B., Itakura, I. and Woo, S.L.C. (1983) *Nature*, **304**, 230-234.
3. Mullis, K.B. and Faloona, F.A. (1987) *Methods in Enzymol.*, **155**, 35-40.
4. Saiki, R., Scharf, S., Faloona, F.A., Mullis, K.B., Horn, G., Erlich, H. and Arnheim, N. (1985) *Science*, **230**, 1350-1354.
5. Kogan, S., Doherty, M. and Githier, J. (1987) *New Engl. J. Med.*, **317**, 985-990.
6. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science*, **239**, 487-491.
7. Applied Biosystems Nucleic Acid Research News (1987), Issue 4.
8. Norrander, J., Kempe, T. and Messing, J. (1983) *Gene*, **148**, 243-250.
9. Veres, G., Gibbs, R.A., Scherer, S.E. and Caskey, C.T. (1987) *Science*, **237**, 415-417.
10. Gibbs, R.A., Nguyen, P.N., McBride, L.J., Koepf, S.M. and Caskey, C.T. (1989) *Proc. Natl. Acad. Sci. USA*, in press.
11. Landegren, U., Kaiser, R., Sanders, L. and Hood, L. (1988) *Science*, **241**, 1077-1080.
12. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L.E. (1986) *Nature*, **321**, 674-679.
13. Murasugi, A. and Wallace, R.B. (1984) *DNA*, **3**, 269-277.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.

Correction of the UDP-glucuronosyltransferase gene defect in the Gunn rat model of Crigler–Najjar syndrome type I with a chimeric oligonucleotide

BETSY T. KREN*, BHUPESH PARASHAR†, PARAMITA BANDYOPADHYAY*, NAMITA ROY CHOWDHURY†‡, JAYANTA ROY CHOWDHURY†‡, AND CLIFFORD J. STEER*§¶

Departments of *Medicine and ‡Cell Biology, University of Minnesota Medical School, Minneapolis, MN 55455; †Departments of Medicine and Molecular Genetics, and §Marion Bessin Liver Research Center, Albert Einstein College of Medicine, Bronx, NY 10461

Edited by Roscoe O. Brady, National Institutes of Health, Bethesda, MD, and approved July 9, 1999 (received for review May 12, 1999)

ABSTRACT Crigler–Najjar syndrome type I is characterized by unconjugated hyperbilirubinemia resulting from an autosomal recessive inherited deficiency of hepatic UDP-glucuronosyltransferase (UGT) 1A1 activity. The enzyme is essential for glucuronidation and biliary excretion of bilirubin, and its absence can be fatal. The Gunn rat is an excellent animal model of this disease, exhibiting a single guanosine (G) base deletion within the *UGT1A1* gene. The defect results in a frameshift and a premature stop codon, absence of enzyme activity, and hyperbilirubinemia. Here, we show permanent correction of the *UGT1A1* genetic defect in Gunn rat liver with site-specific replacement of the absent G residue at nucleotide 1206 by using an RNA/DNA oligonucleotide designed to promote endogenous repair of genomic DNA. The chimeric oligonucleotide was either complexed with polyethylenimine or encapsulated in anionic liposomes, administered i.v., and targeted to the hepatocyte via the asialoglycoprotein receptor. G insertion was determined by PCR amplification, colony lift hybridizations, restriction endonuclease digestion, and DNA sequencing, and confirmed by genomic Southern blot analysis. DNA repair was specific, efficient, stable throughout the 6-month observation period, and associated with reduction of serum bilirubin levels. Our results indicate that correction of the *UGT1A1* genetic lesion in the Gunn rat restores enzyme expression and bilirubin conjugating activity, with consequent improvement in the metabolic abnormality.

UDP-glucuronosyltransferases (UGTs) are a family of membrane-bound enzymes that catalyze the conjugation of numerous xenobiotics and endogenous substrates with glucuronic acid. Of the known isoforms, only UGT1A1 is physiologically relevant in bilirubin glucuronidation and biliary excretion of this potentially toxic metabolite (1, 2). Crigler–Najjar (CN) syndrome is the inherited deficiency of hepatic UGT1A1 activity and is characterized by elevated serum levels of unconjugated bilirubin (3). Of the two types of CN syndrome, type I is more severe and is characterized by a nearly complete absence of UGT1A1 activity, whereas incomplete deficiency of the enzyme is associated with the less severe type II form (4, 5).

The homozygous Gunn rat, a mutant strain of Wistar rat, is an accurate animal model for CN syndrome type I. Its liver lacks UGT1A1 activity because of the deletion of a single guanosine (G) base in *UGT1A1* that results in a frameshift and a premature stop codon (6, 7). Recombinant adenoviral vectors have been used *in vivo* to correct the hyperbilirubinemia in the Gunn rat with persistent expression of the human bilirubin UGT1A1 enzyme for as long as 2 months (8, 9). Significant progress also has been made in overcoming the

immunogenicity of the adenoviral-based vectors (9–11), but their use requires repeated treatments and immunomodulation of the host to maintain therapeutic levels of UGT1A1.

A novel approach, based on mechanisms of DNA repair (12), was reported to correct single nucleotide mutations in episomal and genomic DNA (13, 14). It was observed that an oligonucleotide (ON) composed of both DNA and RNA exhibited increased pairing efficiency with a genomic DNA target (15, 16). The chimeric RNA/DNA ON was designed for increased stability, resistance to nucleases, and improved localization to genomic target sites (17). In a typical duplex structure, the double-stranded region of the molecule is capped by single-stranded thymidine hairpins. The 5' and 3' ends of the molecule are juxtaposed and sequestered by using a 5-bp GC clamp at the 3' end. The RNA residues are 2'-O-methylated to prevent RNase H degradation as well as to improve the formation of joint molecules (18). The homology segment between the RNA/DNA ON and its genomic target is designed with a single mismatch, which promotes the site-directed genomic alteration by endogenous repair pathways (17, 19).

We have used this technology previously to introduce site-specific missense mutations in genomic DNA in cultured human hepatoma cells (20) and in nonreplicating isolated rat hepatocytes (20, 21). In addition, >40% of the rat hepatic factor IX alleles were mutated *in vivo* by using a nonviral delivery system targeted to the hepatocyte via the asialoglycoprotein receptor (21, 22). Both the genomic and phenotypic changes were stable for more than 1 year in quiescent as well as regenerated livers.

Here, we demonstrate that chimeric RNA/DNA ONs can be used for site-directed insertion of a single G nt in genomic DNA from cultured hepatocytes and intact liver of the Gunn rat. The repair process is dose dependent and associated with restoration of the wild-type *Bst*NI restriction endonuclease site in the *UGT1A1* gene. In addition, the phenotypic change is characterized by the hepatic appearance of UGT1A1 protein, secretion of conjugated bilirubin in bile, and decreased serum bilirubin levels. This strategy of genomic alteration circumvents many of the disadvantages associated with viral vector-mediated gene transfer. Our results suggest that site-directed gene repair offers an attractive alternative to gene augmentation using recombinant viruses or hepatocyte transplantation (23) in the treatment of CN syndrome type I.

MATERIALS AND METHODS

Synthesis of the Chimeric ONs. The chimeric RNA/DNA ONs were obtained from Kimeragen (Newtown, PA). They

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office. Abbreviations: ASF, asialofetuin; CN, Crigler–Najjar; ON, oligonucleotide; PEI, polyethylenimine; UGT, UDP-glucuronosyltransferase. ¶To whom reprint requests should be addressed at: Department of Medicine, Box 36 UMHC, University of Minnesota Medical School, 420 Delaware Street SE, Minneapolis, MN 55455. E-mail: steer001@maroon.tc.umn.edu.

were synthesized by using DNA and 2'-O-methyl RNA phosphoramidite nucleoside monomers as described (13). After deprotection and purification by HPLC, more than 98% of the purified ONs were full length. Fluorescently labeled ONs were synthesized by using a fluorescein-modified deoxynucleotide at the initial 5' position of the all-DNA strand.

Polyethylenimine (PEI) and Liposomal Formulations. The 25-kDa PEI (Fluka) was lactosylated by using sodium cyanoborohydride (Sigma) as described (22). For *in vitro* transfections, the chimeric ONs were combined with PEI at nine equivalents of PEI nitrogen per ON phosphate in 0.15 M NaCl. For *in vivo* delivery, the chimeric ONs were complexed with PEI at an ON phosphate/PEI amine ratio of 1:6 in 5% dextrose (22, 24).

Lipid films of dioleoyl phosphatidylcholine/dioleoyl phosphatidylserine/galactocerebroside (Avanti Polar Lipids) were prepared at a 1:1:0.16 molar ratio, hydrated, and extruded down to 0.5 μ m as described (22). For *in vitro* transfections, 150 μ g of UGT1A1/0.5 ml of 0.15 M NaCl was used to hydrate a 0.5-mg lipid film. For *in vivo* delivery, 600 μ g of ON/ml of 5% dextrose was used to hydrate a 2-mg lipid film. Fluorescently labeled ONs were encapsulated in the anionic liposomes by using the same methods. The encapsulation efficiency of the RNA/DNA ONs was >80%.

Cell Culture and Transfections. Gunn rat hepatocytes immortalized by using the simian virus 40 temperature-sensitive large T-antigen were maintained at the permissive (33°C) temperature in supplemented DMEM (25). Cells were detached by using trypsin-EDTA and replated at a density of 2×10^5 cells per 35-mm Primaria (Becton-Dickinson) dish at the nonpermissive temperature (37°C) 24 h before transfection. Cells were transfected in 1 ml of the same medium supplemented with 2.5 mM CaCl_2 by using a 100- μ l aliquot of transfecting solution containing the chimeric ONs complexed to PEI or vehicle alone. After 18 h, 2 ml of medium was added, and the hepatocytes were maintained for an additional 30 h at 37°C before harvesting by scrapping. For repeat transfections, the medium was removed after 48 h and replaced, and the cells were transferred to 33°C for expansion. One week later, the cells were prepared, transfected, and harvested as outlined above.

Gunn rat hepatocytes were isolated by collagenase perfusion as described (25) and plated at a density of 1×10^6 cells/T25 flask in a chemically defined medium (hepatocyte growth medium, HGM) (26). Cells were transfected with 300 μ l of the liposome-encapsulated chimeric ONs, or vehicle alone, in 3 ml of HGM supplemented with 10% heat-inactivated FBS and 2.5 mM CaCl_2 . Three milliliters of FBS-supplemented HGM was added 18 h after transfection, and the cultures were maintained an additional 30 h at 37°C. Parallel transfections of Gunn rat hepatocytes were done with the fluorescein-labeled chimeric ONs, and the cells were analyzed by confocal microscopy as described (20, 22).

In Vivo Delivery Systems. Male rats (≈ 65 g; Harlan Sprague-Dawley) received 200 μ g of fluorescein-labeled chimeric ONs that were either naked, encapsulated in anionic liposomes, or complexed to PEI in 5% dextrose by tail vein injection. For asialoglycoprotein receptor competition, animals received bolus injections of 5 mg/100 g body weight of asialofetuin (ASF) in 0.15 M NaCl 1 min before and 3 min after injection of the fluorescently labeled ONs (27). Tissue samples were frozen in OCT, and the cryosections were fixed for 10 min with 4% paraformaldehyde (wt/vol) in PBS, pH 7.4. Tissue distribution of the fluorescently labeled ONs was determined by confocal microscopy as described (21).

Gunn rats (≈ 80 g; Harlan Sprague-Dawley) were administered aliquots of 200 μ g of UGT1A1 either complexed to PEI or encapsulated in anionic liposomes, or an equal amount of vehicle alone by tail vein injection in 5% dextrose on 5 consecutive days. Seven days and 4 months postinjection,

random liver tissue samples were removed for DNA isolation. At 6 months, bile samples were collected from the animals as described (10) as well as blood and liver tissue for enzyme activity, DNA, and Western blot analysis.

A separate group of Gunn rats (≈ 200 g) was injected on 5 consecutive days with either vehicle, or the chimeric ONs complexed to PEI or encapsulated in liposomes. A total dose of 3 mg/rat (600 μ g/day $\times 5$) was administered by tail vein injection in 5% dextrose. Rats treated a second time received the same dosing schedule. Blood was drawn under ether anesthesia for serum bilirubin levels and alanine aminotransferase activity (Sigma). Bile samples were collected by bile duct cannulation as described (10).

PCR Amplification, Cloning, and Analysis. Genomic DNA larger than 100 bp was isolated by using the high pure PCR template preparation kit (Boehringer Mannheim). DNA from liver tissue samples was isolated as described (21). PCR amplification (30 cycles of 94°C for 45 s, 55°C for 20 s, and 72°C for 45 s) of a 379-nt region of the rat *UGT1A1* gene using the primers 5'-GGGATTCTCAGAATCTAGACATT-3' (sense) and 5'-GTGTGTGGTATAAATGCTGTAGG-3' (antisense) (28) was performed with 300 ng of the isolated DNA. To rule out PCR artifacts, 1 μ g of UGT1A1 alone, or 300 ng of Gunn rat DNA incubated with up to 1.5 μ g of the UGT1A1 chimeric ON, was subjected to PCR amplification. The amplification products were subcloned into the TA cloning vector pCR 2.1 (Invitrogen), and the ligated material was used to transform frozen competent *Escherichia coli*.

After plating, the colonies were lifted onto Micron Separations MagnaGraph nylon filters, replicated, and processed for hybridization with ^{32}P -end-labeled 17-mer ON probes 1206A (5'-ATGTCCTGAAATGACTG-3') or 1206G (5'-ATGTCCTGGAAATGACT-3'). Hybridizations were performed at 37°C for 24 h and the filters were washed as described (20). Plasmid DNA prepared from colonies hybridizing with 1206A or 1206G was sequenced on an ABI 370A sequencer (Perkin-Elmer) by using the mp13 forward and reverse primers as well as a gene-specific primer 5'-CCCATGGTATTTATGAAGGAATATGC-3' corresponding to nucleotides 1071-1106 of the rat *UGT1A1* cDNA (7). The PCR amplicons from DNA samples isolated after 6 months were subjected to *Bst*NI restriction endonuclease digestion and separated by using 1% agarose gel electrophoresis to distinguish the wild type from the mutant *UGT1A1* Gunn rat gene sequence (29).

Southern and Western Blot Analyses. Genomic DNA from the 6-month liver samples was digested sequentially with *Eco*RI then *Bst*NI, and the fragments were resolved by electrophoresis through a 1% agarose gel. After capillary transfer to nitrocellulose membrane, the blots were hybridized for 24 h at 65°C in $6\times$ SSC containing 1% SDS, $5\times$ Denhardt's, and 200 μ g/ml denatured sonicated fish sperm DNA with ^{32}P -labeled probe corresponding to the 379-nt PCR-amplified fragment of the rat *UGT1A1* gene. After hybridization, the filters were washed in $1\times$ standard saline phosphate/EDTA (0.154 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA; SSPE), 0.5% SDS and then $0.1\times$ SSPE, 0.5% SDS at room temperature and 37°C, respectively, and analyzed by phosphorimaging.

Total homogenate and microsomes were isolated from the flash-frozen liver tissue samples by using the buffers and procedures outlined (30). Protein concentrations were determined with the Bio-Rad protein assay kit. Aliquots of 100 μ g of total or microsomal proteins were separated by using 7.5% SDS/PAGE. After electrophoretic transfer onto nitrocellulose membranes, immunoblots were incubated sequentially with H_2O_2 , 5% milk blocking solution, primary antibody (1:5,000) to rat UGT1A1 (28), and horseradish peroxidase-conjugated goat anti-rabbit IgG secondary antibody. UGT1A1 protein was detected by using the Ultra chemiluminescent system (Pierce).

Serum Bilirubin, UGT1A1 Activity, and HPLC Analysis. Serum bilirubin concentrations were determined in blood drawn from the Gunn rats by using a Sigma diagnostic kit. The UGT1A1 enzyme activity was assayed in digitonin-activated liver homogenates with bilirubin as the acceptor aglycone, as described (1, 2). Bile pigments collected from the cannulated bile ducts of each animal group were evaluated for bilirubin glucuronidation by HPLC analysis as described (11). Authentic pigments were used as standards, and pigments were identified by retention times.

RESULTS

UGT1A1 Correction in Cultured Hepatocytes. We designed the chimeric ONs with the hybrid RNA/DNA strand targeting the nontranscribed DNA sequence of the *UGT1A1* gene (Fig. 1). The sequence of the RNA/DNA molecule was identical to that of the mutant gene with one change. An additional G was placed as the center nt within the stretch of nine DNA residues, flanked on both sides by blocks of modified RNA. The genomic target site corresponded to nucleotide 1206 of the complementary strand of the mutant cDNA (7).

Both the liposomal and PEI delivery systems were targeted to the hepatocyte asialoglycoprotein receptor (22, 31). Gunn rat hepatocytes initially were transfected with the fluorescently labeled ONs at 150, 180, and 300 nM concentrations. There was significant cell uptake and nuclear localization of the labeled chimeric molecules in both the immortalized and primary Gunn rat hepatocytes (data not shown). These cells then were transfected with unlabeled UGT1A1 ONs, which were either complexed to PEI or encapsulated in the anionic liposomes. The frequency of G nt insertion at position 1206 was determined by hybridization of duplicate colony lifts of the PCR-amplified and cloned 379-nt stretch of exon 4 of the rat *UGT1A1* gene (28).

The filter lifts were hybridized with the 32 P-end-labeled ON probes 1206A and 1206G (Fig. 2A). The overall frequency of conversion of the targeted nt was calculated by dividing the number of clones hybridizing with the 1206G probe by the total number of clones hybridizing with both probes. G insertion was observed only in hepatocytes transfected with UGT1A1, and not in cells transfected with vehicle or nonspecific chimeric ONs. Additionally, no hybridization of the 1206G probe occurred in clones derived from DNA isolated from untreated hepatocytes and PCR-amplified in the presence of 0.5–1.5 μ g of the UGT1A1 ON. Nucleotide insertion was dose dependent and was as high as 15.3%. In addition, the frequency of G

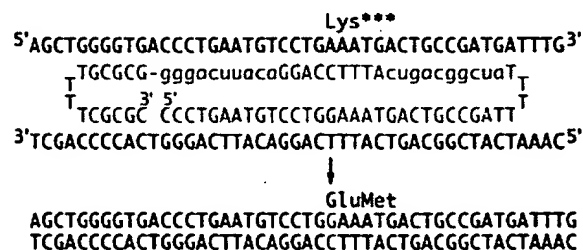


FIG. 1. Targeting strategy to correct the *UGT1A1* frameshift mutation in the Gunn rat. The 2'-O-methylated RNA residues of the targeting RNA/DNA ON (blue) are indicated in lowercase and the DNA residues in capital letters. Blocks of 10 modified RNA residues flank both sides of a 9-residue stretch of DNA, which contains the base change required for correction. The ON sequence is complementary to 28 residues of genomic DNA spanning the site of mutation with the exception of a G base (orange) targeted for position 1206. The cell's endogenous DNA repair process mediates insertion of G at the target site, thereby correcting the frameshift mutation and restoring UGT1A1 activity. The folded double-hairpin structure containing four T residues in each loop, a 5-bp GC clamp, and the modified RNA residues significantly improve resistance to nuclease degradation.

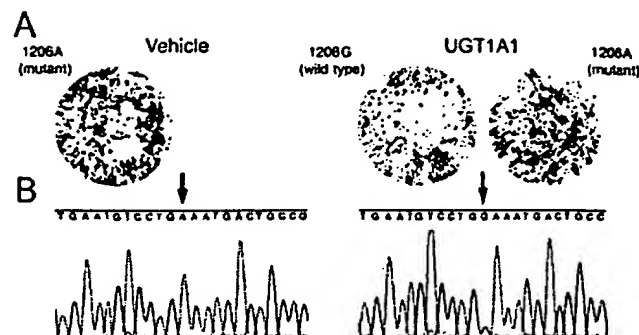


FIG. 2. Filter lift hybridizations and sequence analysis of DNA from isolated hepatocytes. (A) Representative hybridization patterns of duplicate filter lifts of the cloned PCR amplicons with either 32 P-labeled mutant 1206A or wild-type 1206G 17-mer probes. Hepatocytes were transfected with vehicle (Left) or UGT1A1 ON (Right). (B) The nt sequence of plasmid DNA isolated from clones hybridized with probes to mutant 1206A or wild-type 1206G displaying either A (arrow, Left) or G (arrow, Right), respectively.

insertion increased to 23.7% after a second transfection of the immortalized Gunn rat hepatocytes.

We confirmed our results from the filter hybridizations by direct sequencing of at least 12 independent clones of the wild-type and mutant genes (Fig. 2B). The results indicated that colonies hybridizing to only 1206A exhibited the mutant sequence. In contrast, those colonies derived from UGT1A1-transfected Gunn rat hepatocytes hybridizing to the wild-type 1206G ON probe displayed a G at position 1206. The entire 379-nt PCR-amplified region of the *UGT1A1* gene was sequenced for all of the clones and no alterations other than the directed change at the target site was observed. Finally, the start and end points of the 379-nt PCR-amplified genomic DNA samples corresponded exactly to those of the primers used for the amplification process, indicating that the clones sequenced were derived from genomic DNA, rather than nondegraded chimeric ONs.

In Vivo Characterization of the Anionic Liposome and PEI Delivery Systems. The fluorescently labeled ONs, using either PEI or liposomes, were distributed homogeneously throughout the liver as early as 2 h after tail vein injection (Fig. 3). In contrast, there was only minimal uptake in lung, heart, and

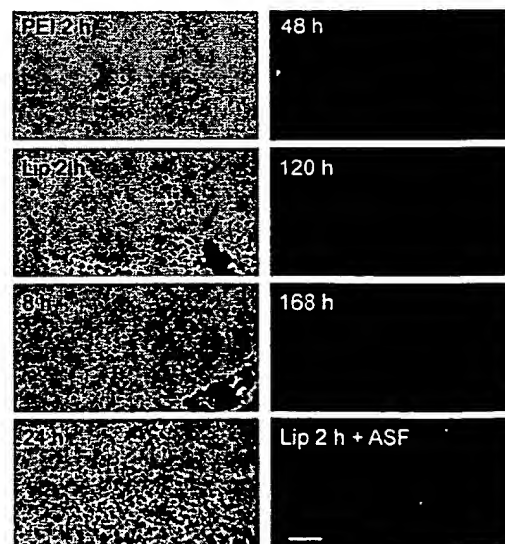


FIG. 3. *In vivo* hepatic distribution of fluorescently labeled ONs. Rats received 200 μ g of 5' fluorescein-labeled chimeric ONs encapsulated in anionic liposomes or complexed with PEI by single bolus tail vein injection. At the indicated times, their livers were processed and examined by confocal microscopy. Lip, liposomes. (Bar = 100 μ m.)

kidney. Coadministration of ASF, which binds avidly to the asialoglycoprotein receptor (27), almost totally inhibited the hepatic uptake of the ONs. This finding was associated with increased levels of the labeled molecules in the other organs. Interestingly, no detectable fluorescence was present in the testis even with coadministration of ASF. In animals injected with naked fluorescein-labeled ONs, there was almost no detectable fluorescence in the liver at 2 h. However, distribution to the other organs was similar to that observed when ASF was coadministered to inhibit liver uptake (data not shown).

We then characterized the time course for hepatic disappearance of the fluorescein label in rats injected with the liposome-encapsulated ONs (Fig. 3). No significant change was observed until 24 h postinjection when the fluorescence began to decline. There was a dramatic decrease throughout the liver by 48 h, and by 120–168 h there was only background fluorescence. Disappearance of the fluorescein label in the other tissues mirrored that observed in the liver. The same pattern of distribution and disappearance was observed when the ONs were complexed to PEI.

In Vivo Correction of the Hepatic *UGT1A1* Mutation. Chimeric ONs complexed with PEI or encapsulated in the anionic liposomes were administered *in vivo* by tail vein injection. Random samples of liver were harvested at 7 days and 4 and 6 months postinjection. Liver DNA was isolated, and the 379-nt sequence spanning the target site was PCR-amplified. Duplicate filter lifts of the transformed colonies were hybridized with the 17-mer ONs to either wild-type 1206G or mutant 1206A. Insertion frequency of G at the genomic target site was $\approx 20\%$ with either delivery system (Fig. 4A) and was unde-

tectable in the control groups. The frequency remained stable at $\approx 20\%$ even when the same livers were analyzed 4 and 6 months postinjection (Table 1). The PCR amplicons from the 6-month samples were subjected to restriction endonuclease digestion with *Bst*NI. Agarose gel analysis indicated partial cleavage at the wild-type *Bst*NI site, whereas DNA from the vehicle controls remained resistant (Fig. 4B, Top). Finally, the 379-nt PCR-amplified DNA fragments were sequenced to confirm G insertion. Amplicons from the *UGT1A1* livers exhibited a mix of wild-type G and mutant A at position 1206 (Fig. 4B, Bottom), whereas the control groups displayed only the mutant A (Middle).

Southern and Western Blot Analyses. DNA was isolated from a variety of liver tissue samples for genomic Southern blot analysis. In fact, DNA isolated from animals that were administered the *UGT1A1* ON showed partial restoration ($\approx 25\%$) of the *Bst*NI restriction site in exon 4 of the *UGT1A1* gene (29) (Fig. 5A). In contrast, the control samples showed no cleavage with *Bst*NI at this site, whereas the wild-type DNA was completely cleaved. The results from the Southern blot analyses were similar for both the PEI and liposomal delivery systems.

Total and microsomal proteins were isolated from liver tissue samples and subjected to Western blot analysis for detection of the 52-kDa *UGT1A1* protein. The results (Fig. 5B) indicated that repair of the *UGT1A1* gene sequence was associated with appearance of the bilirubin-conjugating enzyme. In contrast, there was no detectable *UGT1A1* protein in control samples. The protein was enriched in the microsomal fraction and expressed at 8–15% of wild-type levels, in agreement with the observed enzyme activity in these samples.

Effect of *UGT1A1* Gene Correction on Serum Bilirubin Levels. The serum bilirubin levels of the Gunn rats were monitored after tail vein injection and indicated that a single dosing regimen of the *UGT1A1* molecule, using either PEI or liposomes, resulted in an $\approx 25\%$ decrease in serum bilirubin levels (Fig. 6). In contrast, rats administered vehicle or non-specific ON showed no change, or even an increase in their serum bilirubin levels. A repeat dosing with *UGT1A1* resulted in a further drop in serum bilirubin to $<50\%$ of the pretreatment levels, whereas no significant change was observed in the control rats. Blood studies for routine liver enzymes were performed with both delivery systems, and no changes were detected. Moreover, histologic examination of the livers 6 months after administration indicated that neither PEI, anionic liposomes, nor the chimeric ONs altered liver morphology (data not shown).

Hepatic *UGT1A1* enzyme activity was confirmed by bile duct cannulation and HPLC analysis of bilirubin glucuronidation. In fact, bilirubin mono- and diglucuronides were detected only in those Gunn rats that were administered the *UGT1A1* chimeric ONs (Fig. 7). No significant differences were detected between the PEI and liposomal delivery systems, and in both groups the bilirubin was conjugated primarily as the monoglucuronidated species. Only unconjugated bilirubin was present in the bile of the control Gunn rats.

Table 1. *In vivo* G insertion at nucleotide 1206 of the *UGT1A1* gene in Gunn rat livers

Vehicle	<i>UGT1A1</i> dosage, mg	Insertion, %		
		1 week	4 mos	6 mos
Liposomes	1	20.5 \pm 6.1	17.3 \pm 5.1	19.9 \pm 3.0
PEI	1	23.0 \pm 1.4	19.3 \pm 2.1	20.7 \pm 0.3
PEI control	0	n.d.	n.d.	n.d.

The data represent the mean percentage \pm SD of G insertion from random liver tissue samples determined by filter lift hybridizations as described in *Materials and Methods*. Each treatment group contained at least three animals. n.d., not detectable.

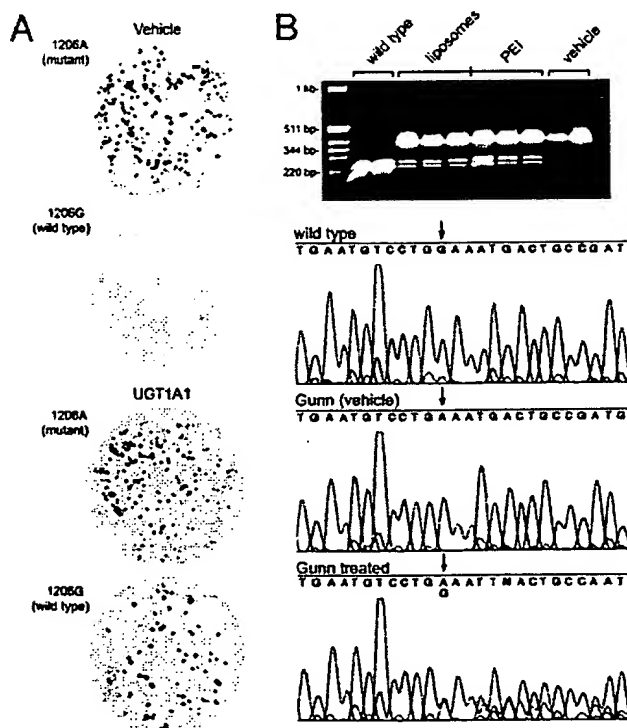


FIG. 4. Filter lift hybridizations, restriction fragment length polymorphism, and sequence analysis of DNA isolated from liver. (A) Hybridization patterns of duplicate filter lifts of the cloned PCR products from liver DNA of Gunn rats 6 months postinjection with vehicle (Upper) or *UGT1A1* ONs (Lower). (B) PCR amplicons were subjected to *Bst*NI restriction enzyme digestion and analyzed by agarose gel electrophoresis and ethidium bromide staining (Top). Direct DNA sequencing of the PCR-amplified *UGT1A1* gene surrounding the targeted G insertion site at position 1206 (arrow) is shown for wild-type (G, top sequence), vehicle (A, middle), and *UGT1A1*-treated Gunn rats (A and G, bottom). The size of the DNA standards is indicated at top left.

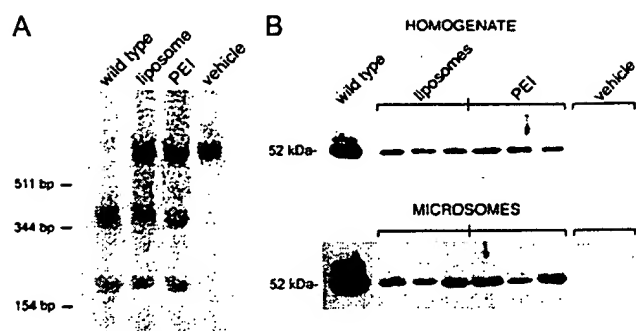


FIG. 5. Southern and Western blot analyses of Gunn rat livers. Liver tissue was harvested for DNA and protein analysis 6 months after *in vivo* administration of the UGT1A1 ONs as described in *Materials and Methods*. (A) Southern blot analysis after sequential digestion of genomic DNA with *Eco*RI and *Bst*NI. (B) Western blot analysis of total liver homogenate and microsomal extracts from the UGT1A1- and vehicle-treated Gunn rats. DNA size markers and protein molecular mass are indicated at left.

DISCUSSION

Chimeric RNA/DNA ONs have been used successfully for single nt substitution in episomal and genomic DNA of replicating cells (13, 14, 20, 32). They also have mediated efficient genomic site-specific nt exchange in isolated nonreplicating as well as quiescent hepatocytes *in vivo* (21, 22). The purpose of this study was to establish whether chimeric ONs could effect site-specific replacement of a G residue to correct the frame-shift mutation in exon 4 of *UGT1A1* in Gunn rats. Our results demonstrate efficient correction of the genetic lesion in both immortalized and primary Gunn rat hepatocytes, as well as the liver *in situ*. In addition, the long-term change together with our previous results with mutation of the rat factor IX gene (22) suggest that correction of the *UGT1A1* gene was permanent.

The genomic insertion of G at the targeted site was not an artifact of PCR amplification, as recently suggested (33). Specifically, neither the control groups nor DNA samples spiked with the UGT1A1 ONs yielded wild-type clones. Also, despite the almost complete hepatic disappearance of the ONs from the liver by 48 h, the frequency of G insertion at 1 week was comparable with that observed 4 and 6 months later in the same livers. Furthermore, hepatic correction of the *UGT1A1*

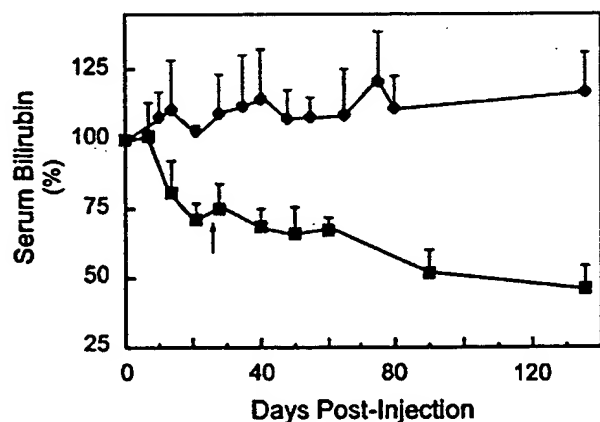


FIG. 6. Effect of *UGT1A1* gene correction on serum bilirubin levels in Gunn rats. Animals were administered UGT1A1 (blue squares) or nonspecific (red circles) ONs complexed to PEI or encapsulated in anionic liposomes as described in *Materials and Methods*. The dosage was repeated for all groups 30 days after the final injection of the first series (arrow). Each data point is the mean \pm SD of 11 animals. There was no significant difference between the PEI and anionic liposome groups. $P < 0.001 \geq 14$ days for UGT1A1 ONs.

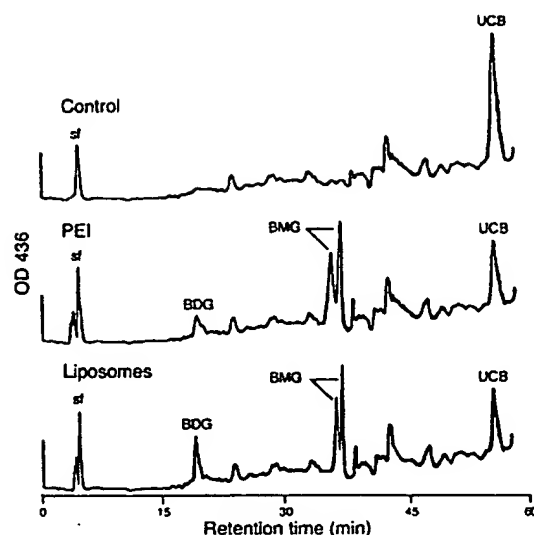


FIG. 7. HPLC analysis of bile pigments from Gunn rat livers. Bile ducts were cannulated and bile collected for HPLC analysis from both PEI- and liposome-treated Gunn rats as described in *Materials and Methods*. BMG, bilirubin monoglucuronide; BDG, bilirubin diglucuronide; UCB, unconjugated bilirubin. The HPLC profiles are representative of four animals in each experimental group.

gene mutation was confirmed by genomic Southern blot analysis, expression of the 52-kDa UGT1A1 enzyme, and a significant reduction in serum bilirubin levels that has been maintained as long as 10 months without additional treatment. In contrast, serum bilirubin remained unchanged, and in some cases increased in control animals. Finally, UGT1A1 enzyme activity was confirmed by the appearance of both mono- and diglucuronidated bilirubin in the Gunn rat bile (1).

The reduction in serum bilirubin levels was gradual and more closely resembled that observed with hepatocyte transplantation (23) rather than whole organ transplantation or overexpression of *UGT1A1* transgenes (8, 10, 11). This finding may be explained by partial correction of the enzyme defect and slow release of bilirubin from the body stores of the Gunn rats, as well as zonal differences in hepatic UGT1A1 expression. The greater proportion of bilirubin monoglucuronide relative to the diglucuronide in bile is also reminiscent of partial bilirubin UGT deficiency states, including CN syndrome type II, Gilbert syndrome, and heterozygous Gunn rats (34, 35). The presence of a higher concentration of bilirubin relative to the number of UGT molecules favors the generation of bilirubin monoglucuronide over the formation of the diglucuronide (36). Also, even with partial gene correction, increased enzyme expression could be achieved by transcriptional induction of *UGT1A1* with several different agents, including phenobarbital (5).

Based on the *in vivo* fluorescent studies, we estimate that $\approx 100,000$ ONs were delivered to hepatocyte nuclei with each tail vein injection. If both alleles are equally amenable to gene repair and only a minority of them are repaired, it is more probable that a single allele would be corrected than both alleles in a given cell. Consistent with this notion, it was reported recently that repair of the tyrosine missense mutation in albino melanocytes appears to occur in a single allele in clonal isolates passaged as many as 10 times (32). This occurrence could be important in some diseases such as α_1 -antitrypsin deficiency, in which a codominant mutant protein may interfere with the function of the wild-type gene product (37).

The incorporation of 2'-O-methylated RNA residues in the structure of the chimeric ON increases the efficiency of nt exchange compared with all-DNA ONs (13, 14, 16). It appears

that the RNA/DNA strand of the duplex is responsible for the initial pairing event, whereas the mismatch within the all-DNA homology strand activates the endogenous DNA repair process (17). In fact, the human recombinase HsRec2 protein, which facilitates homologous pairing (38), significantly increased joint molecule formation between the RNA/DNA ONs and complementary single-stranded DNA compared with the all-DNA ONs (18).

With a novel bacterial test system, the functional capacity of these chimeric molecules to promote targeted nt conversion was shown to require both RecA recombinase and MutS, a mismatch repair enzyme (19). The human MutS homolog *MSH2* also was required for nt conversion in a mammalian cell-free assay system (39). A two-step process was proposed in which RecA mediates strand pairing and formation of a double D-loop, whereas MutS mediates genomic repair (19, 39). In fact, MutS is active in mismatch repair pathways rather than in homologous recombination (40–42). It recently has been reported that the MutS α and MutS β heterodimeric complexes of the mammalian *MSH2* mismatch repair pathway are differentially expressed in cultured cells, and that the *MSH2* protein is involved in modulating their levels (43, 44). Thus, cell lines with varying concentrations of these repair molecules may respond differently to genomic alteration by RNA/DNA ONs (45).

The use of RNA/DNA ONs to correct genetic diseases of the liver offers significant advantages over viral-mediated transgene expression. In particular, it overcomes the random genomic integration associated with certain viral vectors, and the observed immunogenicity and lack of persistent gene expression in others. However, the approach does require the use of ONs that are designed specifically to each genetic mutation. The percent decrease in serum bilirubin levels achieved in this study would be sufficient to convert potentially lethal CN syndrome type I to a manageable CN syndrome type II phenotype. Additionally, the cumulative effect of the repeated treatments coupled with the ability of this technology to induce site-specific nt alteration of genomic DNA without selection offers a potentially powerful technique for both *ex vivo* and *in vivo* gene therapy.

We acknowledge the expert technical assistance of Ms. Xiaoming Ma, Ms. Cheryle Linehan-Stieers, and Mr. Stefan Kren. This work was supported by a grant (to C.J.S.) from Kimeragen, Inc., Newtown, PA; and by Grants RO1-DK 39137 (to N.R.C.), RO1-DK 46057 (to J.R.C.), and P30-DK 41296 (Liver Research Core Center, Albert Einstein College of Medicine) from the National Institutes of Health.

- Roy Chowdhury, N., Gross, F., Moscioni, A. D., Kram, M., Arias, I. M. & Roy Chowdhury, J. (1987) *J. Clin. Invest.* **79**, 327–334.
- Bosma, P. J., Seppen, J., Goldhoorn, B., Bakker, C., Oude Elferink, R. P. J., Roy Chowdhury, J., Roy Chowdhury, N. & Jansen, P. L. M. (1994) *J. Biol. Chem.* **269**, 17960–17964.
- Crigler, J. F. & Najjar, V. A. (1952) *Pediatrics* **10**, 169–172.
- Arias, I. M. (1962) *J. Clin. Invest.* **41**, 2233–2245.
- Roy Chowdhury, J., Wolkoff, A. W., Roy Chowdhury, N. & Arias, I. M. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw-Hill, New York), pp. 2161–2208.
- Iyanagi, T., Watanabe, T. & Uchiyama, Y. (1989) *J. Biol. Chem.* **264**, 21302–21307.
- Roy-Chowdhury, J., Huang, T., Kesari, K., Lederstein, M., Arias, I. M. & Roy-Chowdhury, N. (1991) *J. Biol. Chem.* **266**, 18294–18298.
- Askari, F. K., Hitomi, Y., Mao, M. & Wilson, J. M. (1996) *Gene Ther.* **3**, 381–388.
- Li, Q., Murphree, S. S., Willer, S. S., Bolli, R. & French, B. A. (1998) *Hum. Gene Ther.* **9**, 497–505.
- Takahashi, M., Ilan, Y., Roy Chowdhury, N., Guilda, J., Horwitz, M. & Roy Chowdhury, J. (1996) *J. Biol. Chem.* **271**, 26536–26542.
- Ilan, Y., Prakash, R., Davidson, A., Jona, V., Droguett, G., Horwitz, M. S., Roy Chowdhury, N. & Roy Chowdhury, J. (1997) *J. Clin. Invest.* **99**, 1098–1106.
- Kmieciak, E. B., Cole, A. & Holloman, W. K. (1994) *Mol. Cell. Biol.* **14**, 7163–7172.
- Yoon, K., Cole-Strauss, A. & Kmiec, E. B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 2071–2076.
- Cole-Strauss, A., Yoon, K., Xiang, Y., Byrne, B. C., Rice, M. C., Gryn, J., Holloman, W. K. & Kmiec, E. B. (1996) *Science* **273**, 1386–1389.
- Kotani, H. & Kmiec, E. B. (1994) *Mol. Cell. Biol.* **14**, 6097–6106.
- Kotani, H. & Kmiec, E. B. (1994) *Mol. Cell. Biol.* **14**, 1949–1955.
- Kmieciak, E. B., Kren, B. T. & Steer, C. J. (1999) in *The Development of Human Gene Therapy*, ed. Friedmann, T. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 643–670.
- Havre, P. A. & Kmiec, E. B. (1998) *Mol. Gen. Genet.* **258**, 580–586.
- Kren, B. T., Metz, R., Kumar, R. & Steer, C. J. (1999) *Sem. Liver Dis.* **19**, 93–104.
- Kren, B. T., Cole-Strauss, A., Kmiec, E. B. & Steer, C. J. (1997) *Hepatology* **25**, 1462–1468.
- Kren, B. T., Bandyopadhyay, P. & Steer, C. J. (1998) *Nat. Med.* **4**, 285–290.
- Bandyopadhyay, P., Ma, X., Linehan-Stieers, C., Kren, B. T. & Steer, C. J. (1999) *J. Biol. Chem.* **274**, 10163–10172.
- Fox, I. J., Roy Chowdhury, J., Kaufman, S. S., Goertzen, T. C., Roy Chowdhury, N., Warkentin, P. I., Dorko, K., Sauter, B. V. & Strom, S. C. (1998) *N. Engl. J. Med.* **338**, 1422–1426.
- Abdallah, B., Hassan, A., Benoist, C., Goula, D., Behr, J. P. & Demeneix, B. A. (1996) *Hum. Gene Ther.* **7**, 1947–1954.
- Fox, I. J., Roy Chowdhury, N., Gupta, S., Kondapalli, R., Schilsky, M. L., Stockert, R. J. & Roy Chowdhury, J. (1995) *Hepatology* **21**, 837–846.
- Block, G. D., Locker, J., Bowen, W. C., Petersen, B. E., Katyal, S., Strom, S. C., Riley, T., Howard, T. A. & Michalopoulos, G. K. (1996) *J. Cell Biol.* **132**, 1133–1149.
- Hara, T., Aramaki, Y., Tsuchiya, S., Hosoi, K. & Okada, A. (1987) *Biopharmaceutics. Drug Dispos.* **8**, 327–339.
- Ilan, Y., Roy Chowdhury, N., Prakash, R., Jona, V., Attavar, P., Guha, C., Tada, K. & Roy Chowdhury, J. (1997) *Transplantation* **64**, 8–13.
- Huang, T.-J., Roy Chowdhury, J., Lahiri, P., Yerneni, P. C., Bommineni, V. R., Arias, I. M. & Roy Chowdhury, N. (1992) *Hepatology* **16**, 756–762.
- Liu, L., Jang, G.-F., Farnsworth, C. C., Yokoyama, K., Glomset, J. A. & Gelb, M. H. (1995) *Methods Enzymol.* **250**, 189–206.
- Bandyopadhyay, P., Kren, B. T., Ma, X. & Steer, C. J. (1998) *BioTechniques* **25**, 282–292.
- Alexeev, V. & Yoon, K. (1998) *Nat. Biotech.* **16**, 1343–1346.
- Zhang, Z., Eriksson, M., Falk, G., Graff, C., Presnell, S. C., Read, M. S., Nichols, T. C., Blombäck, M. & Anvret, M. (1998) *Antisense Nucleic Acid Drug Dev.* **8**, 531–536.
- Blanckaert, N. (1980) *Biochem. J.* **185**, 115–128.
- Favery, J., Blanckaert, N., Heirwegh, K. P. M., Préaux, A.-M. & Berthelot, P. (1977) *J. Clin. Invest.* **60**, 970–979.
- Gordon, E. R. & Goresky, C. A. (1980) *Can. J. Biochem.* **58**, 1302–1310.
- Teckman, J. H., Qu, D. & Perlmutter, D. H. (1996) *Hepatology* **24**, 1504–1516.
- Rice, M. C., Smith, S. T., Bullrich, F., Havre, P. & Kmiec, E. B. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 7417–7422.
- Cole-Strauss, A., Gamper, H., Holloman, W. K., Muñoz, M., Cheng, N. & Kmiec, E. B. (1999) *Nucleic Acids Res.* **27**, 1323–1330.
- Yáñez, R. J. & Porter, A. C. G. (1998) *Gene Ther.* **5**, 149–159.
- Kolodner, R. D. (1995) *Trends Biochem. Sci.* **20**, 397–401.
- Modrich, P. (1997) *J. Biol. Chem.* **272**, 24727–24730.
- Genschel, J., Littman, S. J., Drummond, J. T. & Modrich, P. (1998) *J. Biol. Chem.* **273**, 19895–19901.
- Umar, A., Risinger, J. I., Glaab, W. E., Tindall, K. R., Barrett, J. C. & Kunkel, T. A. (1998) *Genetics* **148**, 1637–1646.
- Santana, E., Peritz, A. E., Iyer, S., Uitto, J. & Yoon, K. (1998) *J. Invest. Dermatol.* **111**, 1172–1177.

- Yeasts, A. H. Rose and J. S. Harrison, Eds. (Academic Press, New York, 1969), vol. 1, pp. 385-460.
6. M. S. Esposito, *Proc. Nat. Acad. Sci. U.S.A.* 75, 4436 (1978).
 7. H. Roman, *Carlsberg Res. Commun.* 45, 211 (1980).
 8. F. Fabre, *Nature* 272, 795 (1978).
 9. R. E. Malone, J. E. Golini, M. S. Esposito, *Curr. Gen.* 1, 241 (1980).
 10. M. Meselson and C. Radding, *Proc. Natl. Acad. Sci. U.S.A.* 72, 358 (1975).
 11. C. Bruschi and M. Esposito, *ibid.* 80, 7566 (1983).
 12. R. Holliday, *Genet. Res.* 5, 282 (1964).
 13. J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, *Cell* 33, 25 (1983).
 14. M. Esposito and J. Wagstaff, in *The Molecular Biology of the Yeast Saccharomyces*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), vol. 1, pp. 341-370.
 15. T. L. Orr-Weaver and J. W. Szostak, *Microbiol. Rev.* 49, 33 (1985).
 16. A. Gaudet and M. Fitzgerald-Hayes, *Mol. Cell. Biol.* 7, 68 (1987).
 17. D. Koshland, L. Rutledge, M. Fitzgerald-Hayes, L. H. Hartwell, *Cell* 48, 801 (1987).
 18. S. Fogel, R. K. Mortimer, K. Lusnak, in *The Molecular Biology of the Yeast Saccharomyces*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), vol. 1, pp. 289-339.
 19. J. D. Boeke, F. Lacroute, G. R. Fink, *Mol. Gen. Genet.* 197, 345 (1984).
 20. S. Jinks-Robertson and T. D. Petes, *Genetics* 114, 731 (1986).
 21. M. Lichten, R. H. Borts, J. E. Haber, *ibid.* 115, 233 (1987).
 22. F. Sherman, G. R. Fink, J. B. Hicks, *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983).
 23. H. Ito, Y. Fukuda, K. Murata, A. Kimura, *J. Bacteriol.* 153, 163 (1983).
 24. R. J. Rothstein, *Methods Enzymol.* 101, 202 (1983).
 25. L. Clarke and J. Carbon, *Nature* 305, 23 (1983).
 26. E. M. Southern, *J. Mol. Biol.* 98, 503 (1975).
 27. P. A. Brown and J. W. Szostak, *Methods Enzymol.* 101, 278 (1983).
 28. We thank M. Fitzgerald-Hayes for providing the CEN3 probes, J. Szostak for providing for the TCN⁺ gene on plasmid pSZ414 and S. Jinks-Robertson for providing plasmid pSR7; Nina Brown for technical assistance; S. Picologlou, and D. Morrison for helpful comments about the manuscript; and W. O. Godtfredsen and Leo Pharmaceutical Products for the gift of trichodermin. Supported by NIH grant GM24110 from the National Institute of General Medical Sciences.

25 March 1988; accepted 13 June 1988

A Ligase-Mediated Gene Detection Technique

ULF LANDEGREN, ROBERT KAISER, JANE SANDERS, LEROY HOOD

An assay for the presence of given DNA sequences has been developed, based on the ability of two oligonucleotides to anneal immediately adjacent to each other on a complementary target DNA molecule. The two oligonucleotides are then joined covalently by the action of a DNA ligase, provided that the nucleotides at the junction are correctly base-paired. Thus single nucleotide substitutions can be distinguished. This strategy permits the rapid and standardized identification of single-copy gene sequences in genomic DNA.

DNA ANALYSIS IS ATTAINING increasing importance for the diagnosis of disease caused by single-gene defects as well as for the detection of infectious organisms (1). Moreover, a number of genes, predominantly those encoded in the major histocompatibility complex, have been found to be associated with an increased susceptibility to a variety of disease states (2). Of a total of approximately 2000 defined human genetic loci (3), approximately 100 have currently been studied at the DNA level for their role in genetic disease (4). A number of genetic diseases are caused by alleles present in the population at relatively high frequencies, perhaps because of selective advantages to the heterozygous carriers (5). The ongoing characterization of disease-causing or disease-associated gene sequences makes large-scale screening for carrier status and genetic counseling a possibility. It may also sharpen the diagnostic accuracy for diseases such as autoimmune conditions where the susceptibility may be influenced by defined alleles. Such prospects are currently limited by the cumbersome

nature of the available DNA detection methods.

A majority of polymorphisms in the human genome are caused by point mutations that involve one or a few nucleotides. Current DNA analysis procedures capable of detecting the substitution of a single nucleotide are based on differential denaturation of mismatched probes as in allele-specific oligonucleotide hybridization (6) or denaturing gradient gel electrophoresis (7). Alternatively, the sequence of interest can be investigated for polymorphisms that affect the recognition by a restriction enzyme (8) or that will allow ribonuclease A (RNase A) to cleave at mismatched nucleotides of an RNA probe hybridized to a target DNA molecule (9). Although denaturing gradient gel or RNase A can survey long stretches of DNA for mismatched nucleotides, they are estimated to detect only about half of all mutations that involve single nucleotides (7, 9). Similarly, less than half of all point mutations give rise to gain or loss of a restriction enzyme cleavage site (10). The only existing technique capable of identifying any single

nucleotide difference, short of DNA sequence analysis, is allele-specific oligonucleotide hybridization. This technique involves immobilizing separated (6) or enzymatically amplified (11) fragments of target DNA, hybridizing with oligonucleotide probes, and washing under carefully controlled conditions to discriminate single nucleotide mismatches.

We have devised a strategy that permits the facile distinction of known sequence variants differing by as little as a single nucleotide. The approach combines the ability of oligonucleotides to hybridize to the sequence of interest and the potential of a DNA-specific enzyme, T4 DNA ligase, to distinguish mismatched nucleotides in a DNA double helix (Fig. 1). Two oligonucleotide probes are permitted to hybridize to the denatured target DNA such that the 3' end of one oligonucleotide is immediately adjacent to the 5' end of the other. The ligase can then join the two juxtaposed oligonucleotides by the formation of a phosphodiester bond, provided that the nucleotides at the junction are correctly base-paired with the target strand. The ligation event thus positively identifies sequences complementary to the two oligonucleotides. A heterozygous sample is therefore scored as positive for both alleles. The joining of the oligonucleotides may be conveniently demonstrated, for instance, by labeling one of the oligonucleotides with biotin and the other one with ³²P. After the ligation reaction, the biotinylated oligonucleotides are allowed to bind to streptavidin immobilized on a solid support. Radioactive oligonucleotides that have become ligated to biotinylated oligonucleotides remain on the support after washing and are detected by autoradiography.

The gene encoding human β globin was selected as a model system to test the technique. There are two relatively frequent alleles, β^S and β^C , each differing from the normal allele, β^A , by a single nucleotide substitution in positions 2 and 1, respectively, of codon six (Figs. 2 and 3) (12). Subjects homozygous for the β^S allele suffer from sickle cell anemia. Moreover an increased risk of sudden death during exertion has been observed among individuals heterozygous for β^S (13).

The ligase-mediated gene detection procedure was used to distinguish β^A and β^S genes in equivalent amounts of DNA present in cells, in cloned DNA, and in genomic DNA (Fig. 2). One of two synthetic oligonucleotides (B131 or B132), specific for each of the alleles, was used in conjunction

with another oligonucleotide (P133) hybridizing immediately 3' to either of the other two oligonucleotides on the target DNA strand. All of the synthetic oligomers used in this study are 20 nucleotides long. The ability of T4 DNA ligase to join the variable, 3' nucleotide of the allele-specific oligonucleotides to the 5' terminus of the invariant oligonucleotide was assessed by capturing any ligated product on streptavidin-agarose beads. The beads were filtered and washed to remove unbound oligonucleotides, and then the filter with trapped beads was exposed to x-ray film. The 10^6 nucleated cells used for one assay were obtained from ~0.5 ml of blood. The cells were used in the assay without DNA purification, by first making the DNA accessible for the ligase-mediated analysis by sequential additions of a nonionic detergent (Triton X-100) and a protease (trypsin). The DNA was denatured with alkali and then soybean trypsin inhibitor was added to prevent proteolysis of the added ligase.

The described ligation reactions were performed at 37°C, ~25 K below the melting temperature of the hybridized oligonucleotides, permitting the use of standardized assay conditions independent of the particular sequence investigated. The observed specificity is a consequence of the requirement for the simultaneous hybridization of both oligonucleotides in a precisely juxtaposed position. Although both oligonucleotides are likely to hybridize to numerous sequences in the DNA sample, they are unlikely to do so in the appropriate head-to-tail fashion except where the proper target sequence is present. In addition, we have found that the ligation reaction requires that the two terminal nucleotides on either side of the junction of the two oligonucleotides be engaged in correct base-pairing. This requirement further suppresses incorrect ligation events.

To determine whether any type of single nucleotide mismatch could be distinguished from correct base-pairing with the present method, we used four synthetic target molecules representing a segment of the β -globin gene, each with a different nucleotide in the first position of the sixth codon. Two of the sequences are derived from the β^A and β^C alleles of the β -globin gene. The other two sequences represent the other possible nucleotides occupying the variant position. Four pairs of oligonucleotides were designed to specifically identify one of the target molecules. Four oligonucleotide probes, each with a different nucleotide in the 3' terminal position and complementary to one of the target molecules, were separately assayed for their ability to be ligated to an invariant oligonucleotide that hybrid-

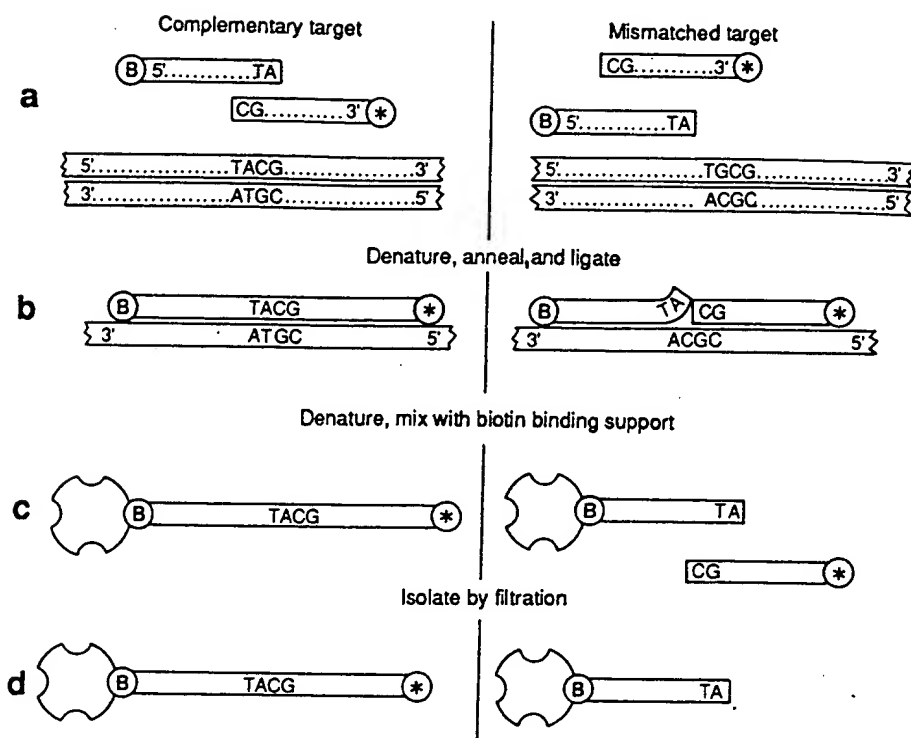
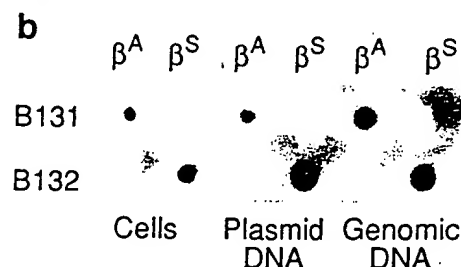


Fig. 1. A diagram depicting gene detection through the ligation of hybridized oligonucleotide probes. Target DNA is denatured and mixed with oligonucleotides and ligase. The ligase joins pairs of oligonucleotides annealed head to tail if they are correctly base-paired at the junction. Radioactively labeled oligonucleotides (*) are immobilized and detected by autoradiography only if ligated to biotinylated oligonucleotides (B) that can be bound to streptavidin on a solid support.

a		MetValHisLeuThrProG1		uGluLysSerAlaValThr	
B131	(β^A)	5'	B	ATGGTGCACCTGACTCCTGA	pGGAGAAGTCTGCCGTTACTG 3' P133
B132	(β^S)	5'	B	-----T	
		Va		1	

Fig. 2. (a) Nucleotide sequence and corresponding translated sequence of the oligonucleotides used in the analysis described in (b). (b) Analysis for the presence of the globin β^A or β^S allele in samples containing equal numbers of copies of β -globin alleles present in nucleated cells, in cloned DNA, and in genomic DNA. Two $\times 10^6$ linearized plasmid molecules containing the β^A or β^S allele of the human globin genes were added to individual microtiter wells containing 10 μ g of salmon sperm DNA in 4 μ l of water (19). The microtiter plates were centrifuged and the supernatants removed. To the resuspended cell pellet was added 1 μ l of 10% Triton X-100 and 1 μ l of trypsin at 2 μ g/ μ l. The samples were incubated at 37°C for 30 min and were denatured with alkali as above. The pH was neutralized and 1 μ l of soybean trypsin inhibitor (Sigma, 10 μ g/ μ l) was added. Each well received 140 fmol of biotinylated oligonucleotides B131 or B132 (20), specific for the globin β^A and β^S genes, respectively, and 1.4 fmol of oligonucleotide P133, 5' end-labeled with [γ - 32 P]adenosine triphosphate (ATP) and polynucleotide kinase to a specific activity of 5×10^8 Cerenkov cpm/ μ g and purified over a Nensorb column (Du Pont Biotechnology Systems). T4 DNA ligase (0.05 Weiss unit, Collaborative Research) was added in 2 μ l of 5 \times ligase buffer to a final volume of 10 μ l containing 50 mM tris-HCl (pH 7.5), 10 mM MgCl₂, 150 mM NaCl (including 50 mM added during denaturation), 1 mM spermidine, 1 mM ATP, 5 mM dithiothreitol, and 100 ng of bovine serum albumin per microliter. The reagents were mixed by briefly centrifuging the microtiter plates before incubating at 37°C and 100% humidity for 5 hours. The ligated oligonucleotides were denatured by the addition of 1 μ l of 1.1M NaOH and incubated for 10 min at 37°C. After the incubation, 1 μ l of 1.1M HCl and 2 μ l of 10% SDS were added. Three microliters of a 15% (v/v) suspension of streptavidin-coated agarose beads (Bethesda Research Laboratories) was then added, and the plate was incubated on a shaking platform at room temperature for 5 min. The contents of the wells were transferred to a dot blot manifold (Schleicher and Schuell) with a Whatman filter paper no. 4. In order to reduce nonspecific binding of the labeled oligonucleotides, the filter papers had been boiled and the beads diluted in 0.5% (v/v) dry nonfat milk, 1% SDS, and salmon sperm DNA (100 μ g/ml). The beads (21) were washed under suction in the manifold with 3 ml of 1% SDS and 1 ml of 0.1M NaOH per sample, with a 96-tip dispenser (Vaccu-pette/96, Culture Tek). The filters were wrapped in plastic wrap and autoradiographed for 3 days at -70°C with one enhancing screen (Du Pont).



ized immediately 3' to the first oligonucleotide. These reagents permit studying the effect on ligation by any of the 16 possible base pairs, the 4 correct Watson-Crick pairs and 12 mismatched pairs, in an invariant sequence context. Under the appropriate conditions, only nucleotides engaged in correct base-pairing were efficiently joined by ligation (Fig. 3). Parameters that affected the nucleotide specificity were the salt concentration and the amount of enzyme added relative to the DNA concentration. Higher salt concentration and lesser amounts of enzyme than those found to be optimal for discrimination resulted in loss of signal. The above experiment cannot exclude the possibility that the identification of mismatched nucleotides may be influenced by the surrounding sequence, although we have not yet encountered any evidence for such effects.

Although autoradiographic techniques are relatively simple to implement, a gene detection assay based on the use of fluorescent rather than radioactive probes would have the advantages of safe handling, more stable reagents, and rapid access to the results, and would allow for multicolor analysis by using fluorophores with different emission spectra. In general, conventional organic fluorophores are less sensitive labels than ^{32}P . Thus we increased the amount of target DNA before the detection assay with the polymerase chain reaction (14). With

this procedure a segment of DNA can be exponentially amplified by repeated cycles of enzymatic synthesis of new strands from two oligonucleotide primers, one with a sequence derived upstream and the other in the opposite orientation downstream of the segment of interest. Genomic DNA was obtained from three human cell lines, MOLT-4, which is homozygous for the β^A -globin allele; SC-1, homozygous for the β^S allele; and GM2064, in which the β -globin locus has been deleted (15). The appropriate segment of the β -globin gene was amplified in 25 cycles from 1 μg of genomic DNA from each cell line. We used 3- μl aliquots, equivalent to 24 ng of genomic DNA for the assay. Two oligonucleotides, specific for the β^A and β^S alleles and differentially 5'-labeled with one of two fluorophores, were present at equal concentrations. The amount of each of these oligonucleotides that became ligated to a third oligonucleotide hybridizing downstream of the other two was determined by separating the reaction products on an 8% polyacrylamide gel and analyzing the band migrating as a 40-nucleotide oligomer (the size of two ligated oligonucleotides) for the relative contribution by the two different fluorophores [model 370A DNA sequencer, Applied Biosystems, Foster City, California (16)]. No signal was observed when the β -globin gene had been deleted in the cell from which the DNA was

obtained, whereas only the correct fluorophore-labeled oligonucleotide was ligated when the cells harbored the β^A or β^S alleles

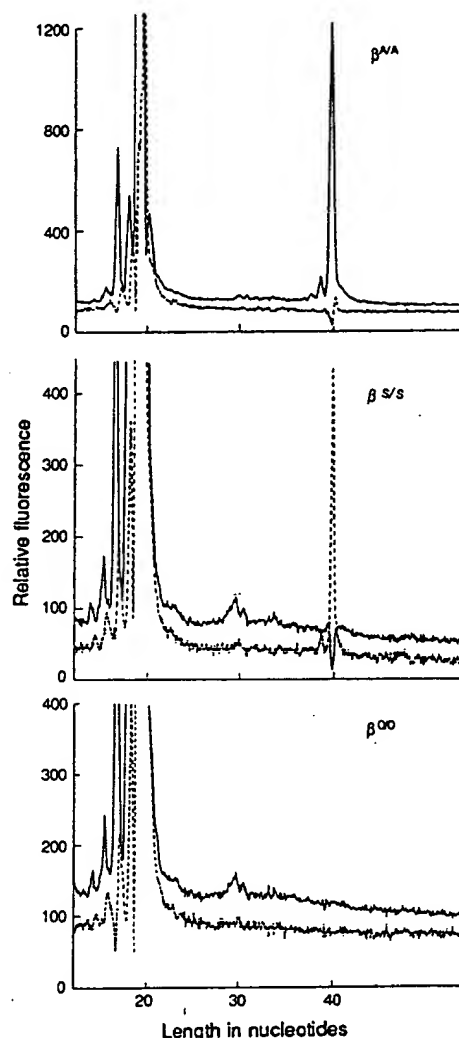
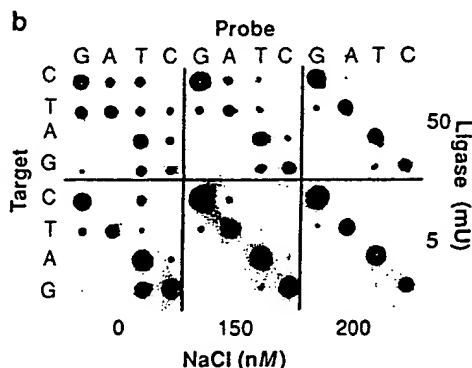


Fig. 4. Demonstration of the presence of the β^A and β^S alleles of the β -globin gene in amplified genomic DNA by probes labeled with fluorescent dyes. A 120-bp segment of the β -globin gene was amplified with the polymerase chain reaction as described (16) in 25 cycles starting with 1 μg of genomic DNA from the cell lines MOLT-4, SC-1, and GM2064 (β^A , β^S , and β^0 , respectively) in 100 μl . Three microliters of each amplified sample was added to an Eppendorf tube, denatured by alkali, neutralized, and incubated with 14 fmol each of oligonucleotide 131 labeled with carboxy-fluorescein (Molecular Probes) (CF131) (—) and oligonucleotide 132 labeled with carboxy-2',7'-dimethoxy-4',6'-dichlorofluorescein (CD132) (---), and 14 fmol of nonradioactively 5' phosphorylated oligonucleotide P133 (for sequences, see Fig. 2). The reaction conditions were essentially as described in Fig. 2, but 0.5 Weiss unit of T4 DNA ligase was added to each assay. At the end of the 3-hour incubation, the samples were ethanol precipitated, taken up in 50% formamide, and loaded on a sequencing gel in an ABI 370A automated DNA sequencer. The fluorescence signal was processed to distinguish the partially overlapping emission spectra of the two fluorophores and to determine the relative contribution of each fluorophore to the signal.

a

B128	(β^A)	5'	B	CATGGTGCACCTGACTCCTG	pAGGAGAAGTCTGCCGTTACT	3'	P129
B134	(β^C)	5'	B	-----A			
B136		5'	B	-----T			
B137		5'	B	-----C			
172	(β^A)	3'		GTACCACGTGGACTGAGGACTCCTCTTCAGACGGCAATGA		5'	
138	(β^C)	3'		-----T-----		5'	
139		3'		-----A-----		5'	
140		3'		-----G-----		5'	

Fig. 3. (a) Nucleotide sequence of the oligonucleotides used in the analysis described in (b). (b) Correct identification of four target molecules, differing by single-nucleotide substitutions in one position. Letters refer to the variable nucleotides in the probe and target sequences. As target molecules, 40-nucleotide oligomers, derived from the β -globin gene sequence, were synthesized. The oligonucleotides 172, 138, 139, and 140 are of identical sequence except in a central position where each target molecule includes a different nucleotide. Four 20-nucleotide biotinylated oligomers, B128, B134, B136, and B137, differing only in their 3' nucleotide position, were designed to hybridize to the 3' half of the target molecules such that the variant position of the probe reagents corresponds to that of the target molecules. Each of the biotinylated oligonucleotides was used in conjunction with oligonucleotide P129, 5' end-labeled with ^{32}P and hybridizing immediately 3' to the biotinylated probes on the target strands. The assays were performed essentially as described in the legend to Fig. 2, but 2×10^8 copies of one of the target molecules were added to each well with 10 μg of salmon sperm DNA. Each well further received one of the biotinylated oligonucleotides together with oligonucleotide P129. The final NaCl and ligase concentrations were varied as indicated.



(Fig. 4). This strategy could be generalized to the simultaneous analysis of several loci. For each set of two labeled, allele-specific oligonucleotides and one unlabeled, the latter is given a nonhybridizing 3' sequence extension of a unique length. This results in different migration rates for the ligation products, characteristic of each locus.

In contrast to gene detection techniques based on immobilizing the target DNA, such as DNA blots, the hybridization reported here was performed in solution and in a small volume, which reduced the time required for hybridization (17). It also obviated the step of immobilizing the target DNA. Both ligation and binding of the biotinylated oligonucleotides are efficient and rapid steps that should permit quantitative detection of target molecules. In general, there are three rate-limiting steps in gene detection techniques. The first is sample preparation, which can be greatly simplified as demonstrated here. The second is the time required for the probes to anneal to the target sequence. This is a function of the concentration of the probe and can be reduced considerably. The third and most time-consuming step in the present technique is signal detection by autoradiography. A sufficiently sensitive fluorescent detection method (18) should drastically reduce this time, permitting the development of a rapid, automated gene detection procedure.

REFERENCES AND NOTES

1. C. T. Caskey, *Science* 236, 1223 (1987).
2. A. Sveigaard et al., *Immunol. Rev.* 70, 193 (1983); J. L. Tiwari and P. I. Terasaki, *HLA and Disease Associations* (Springer-Verlag, New York, 1985).
3. V. A. McKusick, *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore, ed. 7, 1986).
4. D. N. Cooper and J. Schmidtke, *Lancet* i, 273 (1987).
5. J. I. Rotter and J. M. Diamond, *Nature* 329, 289 (1987).
6. B. J. Conner et al., *Proc. Natl. Acad. Sci. U.S.A.* 80, 278 (1983); N. M. Whiteley, M. W. Hunkapiller, A. N. Glazer, European Patent application 0 185 494 A2 (1986).
7. R. M. Myers et al., *Nature* 313, 495 (1985).
8. R. F. Geever et al., *Proc. Natl. Acad. Sci. U.S.A.* 78, 5081 (1987).
9. R. M. Myers, Z. Larin, T. Maniatis, *Science* 230, 1242 (1985); E. Winter, F. Yamamoto, C. Almoguera, M. Perucho, *Proc. Natl. Acad. Sci. U.S.A.* 82, 7575 (1985).
10. T. Hunkapiller, personal communication.
11. R. K. Saiki, T. L. Bugawan, G. T. Horn, K. B. Mullis, H. A. Erlich, *Nature* 324, 163 (1986).
12. R. M. Winslow and W. F. Anderson, in *The Metabolic Basis of Inheritance*, J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson, J. L. Goldstein, M. S. Brown, Eds. (McGraw-Hill, New York, 1983), pp. 1666-1710.
13. J. A. Kark, D. M. Posey, H. R. Schumacher, C. J. Ruchle, *N. Engl. J. Med.* 317, 781 (1987).
14. R. K. Saiki et al., *Science* 230, 1350 (1985).
15. R. K. Saiki, N. Arnheim, H. A. Erlich, *Biotechnology* 3, 1008 (1985).
16. L. M. Smith et al., *Nature* 321, 674 (1986); C. Connell et al., *BioTechniques* 5, 342 (1987).
17. J. G. Wetmur and N. Davidson, *Mol. Biol.* 31, 349 (1968).

18. E. Soini and H. Kojola, *Clin. Chem.* 29, 65 (1983).
19. Genomic DNA was purified from guanidinium HCl-solubilized cells as described [D. Bowtell, *Anal. Biochem.* 162, 463 (1987)] and resuspended by boiling before adding 7 μ g in 4 μ l of water per assay well. The plasmid and genomic DNA samples were denatured by adding 1 μ l of 0.5M NaOH and incubating for 10 min at 37°C before restoring the pH with 1 μ l of 0.5M HCl. Alternatively, samples of nucleated blood cells were used directly as a source of DNA for the analysis. Cells (10^6), obtained by Ficoll-Hypaque (Pharmacia) flotation, were added in 50 μ l of phosphate-buffered saline to each well.
20. The oligonucleotides were assembled by the phosphoramidite method [S. J. Horvath, J. R. Firca, T. Hunkapiller, M. W. Hunkapiller, L. Hood, *Methods Enzymol.* 154, 314 (1987)] on an Applied Biosystems model 380A DNA synthesizer and purified either by polyacrylamide gel electrophoresis or reversed-phase high-pressure liquid chromatography (HPLC). Biotinylation was performed by reacting a biotin N-hydroxysuccinimide ester (Enzotol, Enzo) with a 5' aminothymidine residue incorporated in the oligonucleotide [L. M. Smith, S. Fung, T. J. Hunkapiller, M. W. Hunkapiller, L. Hood, *Nucleic Acids Res.* 13, 2399 (1985)]. The product was purified by reversed-phase HPLC.
21. The size of the area on which the beads were deposited was reduced by interposing a 3-mm-thick plexiglass disk with conical holes with diameters of 5 mm on the upper surface and 2 mm on the lower.
22. The authors acknowledge a stipend from the Knut and Alice Wallenberg Foundation to U.L. and support from NSF grant BNS 87 14486, Defense Advanced Research Projects Agency grant N00014-86K-0755, Upjohn Company, and Applied Biosystems, Inc. The oligonucleotides were synthesized by S. J. Horvath and the fluorescence data were analyzed by C. Dodd. R. K. Saiki provided plasmids and samples of genomic DNA obtained from cell lines. J. Korenberg and K. Tanaka made available blood samples from sickle cells patients. The N-hydroxysuccinimide ester of carboxy-2',7'-dimethoxy-4',5'-dichlorofluorescein was provided by M. W. Hunkapiller. We acknowledge discussions with B. Korber, B. Popko, A. Kamb, N. Lan, L. Smith, R. Barth, V. A. McKusick, J. Richards, and M. Simon.

11 April 1988; accepted 23 June 1988

Amyloid Protein Precursor Messenger RNAs: Differential Expression in Alzheimer's Disease

M. R. PALMERT, T. E. GOLDE, M. L. COHEN, D. M. KOVACS, R. E. TANZI, J. F. GUSELLA, M. F. USIAK, L. H. YOUNKIN, S. G. YOUNKIN*

In situ hybridization was used to assess total amyloid protein precursor (APP) messenger RNA and the subset of APP mRNA containing the Kunitz protease inhibitor (KPI) insert in 11 Alzheimer's disease (AD) and 7 control brains. In AD, a significant twofold increase was observed in total APP mRNA in nucleus basalis and locus ceruleus neurons but not in hippocampal subicular neurons, neurons of the basis pontis, or occipital cortical neurons. The increase in total APP mRNA in locus ceruleus and nucleus basalis neurons was due exclusively to an increase in APP mRNA lacking the KPI domain. These findings suggest that increased production of APP lacking the KPI domain in nucleus basalis and locus ceruleus neurons may play an important role in the deposition of cerebral amyloid that occurs in AD.

ALZHEIMER'S DISEASE (AD) IS characterized pathologically by large numbers of senile plaques and neurofibrillary tangles throughout the cerebral cortex and hippocampus. Senile plaques consist of clusters of degenerating neurites surrounding an amyloid core composed of 5- to 10-nm fibrils that stain metachromatically with Congo red. In many cases of AD, amyloid fibrils are also found in vessel walls (1). A 4.2-kD polypeptide, referred to as A β or the β protein, has been isolated from the amyloid fibrils found in senile plaques (2) and vessel walls (3) of patients with AD. There is evidence that A β may also be a component of the paired helical filaments found in neurofibrillary tangles (4).

The gene encoding A β , which is located on chromosome 21 (5), produces at least three mRNAs (Fig. 1) referred to as APP₆₉₅, APP₇₅₁, and APP₇₇₀ (6-8). APP₆₉₅, the mRNA that was initially identified (5), en-

codes an amyloid protein precursor (APP), 695 amino acids in length, that includes A β at positions 597 to 638. APP₇₅₁ is identical to APP₆₉₅, except for a 168-nucleotide insert (6-8). This insert, previously referred to as HL124i (7), would introduce 56 amino acids carboxyl terminal to Arg²⁸⁸ and convert Val²⁸⁹ into an isoleucine. APP₇₇₀ is identical to APP₇₅₁, except for a 57-nucleo-

M. R. Palmert, T. E. Golde, M. F. Usiak, L. H. Younkin, S. G. Younkin, Division of Neuropathology, Institute of Pathology and Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, OH 44106.

M. L. Cohen and D. M. Kovacs, Division of Neuropathology, Institute of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106. R. E. Tanzi and J. F. Gusella, Neurogenetics Laboratory, Massachusetts General Hospital and Department of Genetics and the Program in Neuroscience, Harvard Medical School, Boston, MA 02115.

*To whom correspondence should be addressed at the Department of Pharmacology.

A novel method for the parallel analysis of multiple mutations in multiple samples

Uwe Maskos and Edwin M. Southern

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received February 25, 1993; Accepted March 25, 1993

In a novel method for analysing mutations, allele specific oligonucleotides (ASOs) are synthesised in stripes on the surface of a glass plate and single-stranded ^{35}S labelled RNA probes applied in orthogonal stripes. We have tested the approach using the well studied example of the sickle cell mutation in the human β -globin gene.

Detection of sequence variation in DNA has applications in linkage analysis, in the analysis of inherited diseases, in genetic fingerprinting, and in studies of evolution. As full sequence analysis is time consuming and expensive, several methods have been developed to analyse variation directly. Some are based on the hybridisation of short oligonucleotides to the test sequence (1). One advantage of this approach is that it can detect both the mutant and the wild-type sequence in a single analysis. In others the oligonucleotides may be bound to a solid support and probed with the labelled test sequence (2, 3).

In some applications, such as the analysis of a common mutation as may be the case in the haemoglobinopathies (4), there is a need to analyse many samples with a few oligonucleotides; in others, for example in linkage analysis using ASOs instead of RFLPs (5), there is a need to analyse a few samples with many oligonucleotides; and in yet others, for example population screening for mutations in genes such as the *CFTR* gene with many alleles (6), there is need to analyse many samples with many oligonucleotides. This communication describes a versatile approach which can be adapted to any of these applications.

Our method for synthesising oligonucleotides on glass plates (7, 8) was used to produce stripes of ASOs 15 nt long for the A, C and S alleles of the β -globin sickle cell locus. The stripes were 2 mm wide and 150 mm long made using the device shown in Figure 1 on 3 mm thick window glass.

Four different single-stranded RNA probes covering the site of the β -globin mutation were prepared as follows: Carrier and patients' DNAs (a gift from Dr J. Old, John Radcliffe Hospital, Oxford) and wild-type control were amplified using a standard

PCR procedure (25 cycles; 55°C, 2 min; 72°C, 2 min; 94°C, 2 min; Cetus PCR machine), Figure 2, to give a 162 bp product. The 46 nt upstream primer consisted of 20 nt of β -globin sequence and a 26 nt T7 polymerase promoter clamp at the 5' end, and the downstream primer of 20 nt of β -globin sequence and a 26 nt SP6 RNA polymerase clamp to allow separate transcription of either the 'sense' or 'anti-sense' strand.

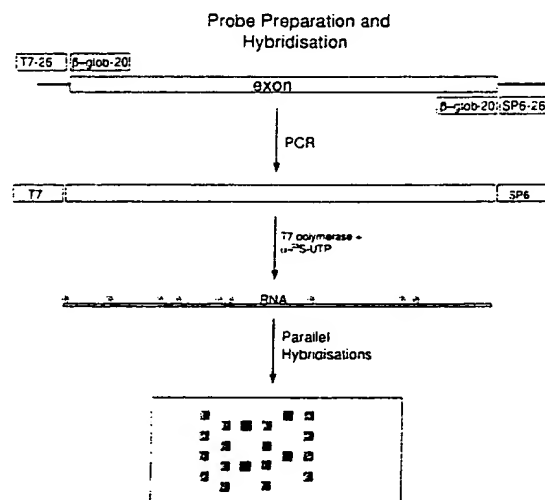


Figure 2. Strategy used to generate a labelled single-stranded RNA probe from genomic DNA. The primers were: upstream primer, TTC TAA TAC GAC TCA CTA TAG GGA GA ACA CAA CTG TGT TCA CTA GC; downstream primer, CTT AAT TAG GTG ACA CTA TAG AAT AG CAA CTT CAT CCA CGT TCA CC. Standard PCR buffer containing 2 mM Mg^{2+} and AmpliTaq polymerase were used.

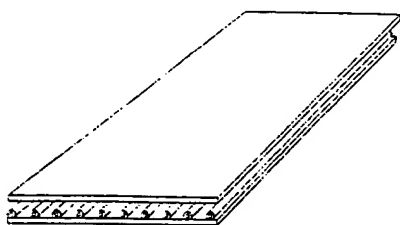


Figure 1. Set-up used to synthesise oligonucleotides in lines on a glass plate.

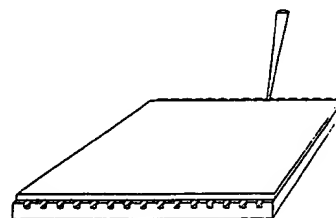


Figure 3. Device used to apply solutions in channels orthogonal to the oligonucleotide lines. Individual channels were 3.5 mm wide.

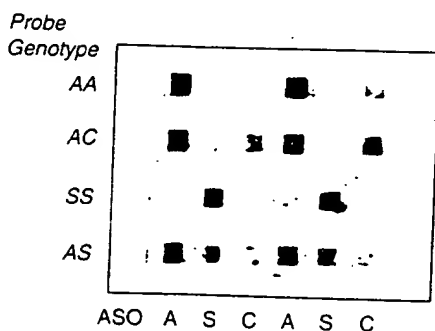


Figure 4. 'Multiplex' analysis of sickle cell mutations. The order of ASO lines is ASC ASC, from left to right. Four single-stranded probes were applied in columns perpendicular to the stripes. The determined genotypes are, from top to bottom, AA, wild-type control; AC, a carrier of the C allele; SS, sickle-cell patient; AS, carrier of the sickle trait.

Transcription was carried out using the Promega Riboprobe kit according to manufacturer's instructions, with 120 μ Ci α - 35 S-UTP (Amersham) and no unlabelled UTP, for 30 min at 37°C. The mixture was spun through a Sephadex G-25 STE column. Equal amounts of radioactive transcript were used in each hybridisation.

Several probes can be analysed simultaneously by applying solutions in stripes across the stripes of oligonucleotides on the surface of the microscope slide. The stripes were formed by putting the plate on the device shown in Figure 3, made from plexiglass. The hybridisation solution (10 μ l, 0.1 M NaCl in TE pH 7.5, containing 0.2% SDS) was run into the line of contact between the plexiglass and the glass by capillary action. Hybridisation was for 2 hrs at room temperature. The plate was rinsed in 0.1 M NaCl, eluted at 43°C for 10 min, and exposed to a PhosphorImager storage phosphor screen overnight, scanned on a Molecular Dynamics PhosphorImager and printed (Figure 4). The results for all individuals are clear and as expected.

The method has several advantages over alternatives. It allows for multiple comparisons to be carried out in a single simple experiment. The number of oligonucleotides that could be synthesised and the number of probes analysed are determined by the size of the glass plate and the width of the stripes. On a 200 mm \times 200 mm plate it should be possible to synthesise 100 different sequences and test probes from 50 different individuals. This level of highly parallel analysis is potentially a lot higher than the method introduced by Erlich *et al.* (9) which involves the hybridisation of one PCR product at a time to oligonucleotides UV crosslinked to strips of membrane.

The manipulations are simple to carry out manually, although the method would lend itself to automation. The analysis is rapid; we have carried out a complete procedure, starting with genomic DNA to having the final result in less than a working day. It is versatile and can be applied to any locus for which there is sufficient information to produce oligonucleotides for test and amplification, e.g. cystic fibrosis testing or scanning for mutations in the *p53* gene. The glass plates, unlike filters, are stable; we have reused them more than thirty times with no loss of performance. The interpretation of the result is straightforward and can readily be automated and quantified.

REFERENCES

- Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979) *Nucleic Acids Res.* 6, 3543-3557.
- Saiki, R.K., Walsh, P.S., Levenson, C.H. and Erlich, H.A. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6230-6234.
- Zhang, Y., Coyne, M.Y., Will, S.G., Levenson, C.H. and Kawasaki, E.S. (1991) *Nucleic Acids Res.* 19, 3929-3933.
- Orkin, S.H. and Kazazian, H.H., Jr (1984) *Annu. Rev. Genet.* 18, 131-171.
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) *Am. J. Hum. Genet.* 32, 314-331.
- Davies, K. (1990) *Nature* 348, 110-111.
- Maskos, U. and Southern, E.M. (1992) *Nucleic Acids Res.* 20, 1679-1684.
- Southern, E.M., Maskos, U. and Elder, J.K. (1992) *Genomics* 13, 1008-1017.
- Erlich, H., Bugawan, T., Begovich, A.B., Scharf, S., Griffith, R., Saiki, R., Higuchi, R. and Walsh, P.S. (1991) *Eur. J. Immunogenet.* 18, 33-55.

References and Notes

1. B. N. Fields, Ed., *Virology* (Raven, New York, 1985); H. R. B. Pelham, *Nature (London)* 272, 469 (1978); *Virology* 97, 256 (1979); M. March, G. Drugeon, C. Benicourt, *ibid.* 119, 193 (1982); E. G. Strauss, C. M. Rice, J. H. Strauss, *ibid.* 133, 92 (1984).
2. Y. Yoshinaka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 82, 1618 (1985).
3. R. Weiss, N. Teich, H. Varmus, J. Coffin Eds., *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), chaps. 4-6.
4. D. E. Schwartz, R. Tizard, W. Gilbert, *Cell* 32, 853 (1983).
5. K. Beemon and E. Hunter, *Proc. Natl. Acad. Sci. U.S.A.* 74, 3302 (1977); A. Purchio, E. Erikson, R. L. Erikson, *ibid.*, p. 4661; S. R. Weiss *et al.*, *Cell* 15, 607 (1978).
6. D. A. Melton *et al.*, *Nucleic Acids Res.* 12, 7035 (1984).
7. H. Oppermann, J. M. Bishop, H. E. Varmus, L. Levintow, *Cell* 12, 993 (1977).
8. J. Ghayeb and M. Inouye, *J. Biol. Chem.* 259, 463 (1984).
9. T. Jacks and H. E. Varmus, unpublished data. RNA was recovered from 100- μ l RRL reaction mixtures with two phenol extractions and subsequent ethanol precipitation in the presence of 2.5M ammonium acetate and used to program either 18- μ l *E. coli* or 25- μ l RRL translation reactions as in Fig. 4.
10. P. Valenzuela, M. Quiroga, J. Zaldivar, P. Gray, W. J. Rutter, in *Animal Virus Genetics*, B. Fields, R. Jaenisch, D. F. Fox, Eds. (Academic Press, New York, 1980), pp. 57-70.
11. J. J. Dunn and F. W. Studier, *J. Mol. Biol.* 166, 477 (1983).
12. J. R. Roth, *Cell* 24, 601 (1981).
13. V. M. Vogt, D. A. Bruckstein, A. P. Bell, *J. Virology* 44, 725 (1982).
14. J. F. Atkins, R. F. Gesteland, B. R. Reid, C. W. Anderson, *Cell* 18, 1119 (1979).
15. R. Weiss and J. Gallant, *Nature (London)* 302, 389 (1983).
16. R. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* 81, 5797 (1984).
17. M. Yarus and R. Thompson, in *Gene Function in Prokaryotes*, J. Beckwith, J. Davies, J. A. Gallant, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), pp. 23-63.
18. T. D. Fox and B. Weiss-Brummer, *Nature (London)* 288, 60 (1980).
19. W. J. Craig, R. G. Cook, W. P. Tate, C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* 82, 3616 (1985).
20. C. W. Rettenmier, R. E. Karess, S. M. Anderson, H. Hanafusa, *J. Virology* 32, 102 (1979).
21. S. Wain-Hobson, P. Sonigo, D. Danos, S. Cole, M. Alizon, *Cell* 40, 9 (1985); C. A. Rattner *et al.*, *Nature (London)* 331, 277 (1985); R. Sanchez-Pescador *et al.*, *Science* 227, 506 (1985).
22. K. Shimotohno *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 82, 3101 (1985).
23. N. Sagata *et al.*, *ibid.*, p. 677.
24. N. R. Rice, R. M. Stevens, A. Burny, R. V. Gliden, *Virology* 142, 357 (1985).
25. J. Clare and P. Farabaugh, *Proc. Natl. Acad. Sci. U.S.A.* 82, 2829 (1985).
26. K. Saigo *et al.*, *Nature (London)* 312, 659 (1985).
27. J. Mellor *et al.*, *ibid.* 313, 243 (1985).
28. V. F. de la Cruz, W. Navelmann, L. Simpson, *J. Biol. Chem.* 259, 15136 (1984).
29. L. Hensgens *et al.*, *Nucleic Acids Res.* 12, 7327 (1984).
30. R. Benne, *Trends Genet.* 1, 117 (1985).
31. W. J. Delorbe *et al.*, *J. Virology* 36, 50 (1980).
32. H. Oppermann, J. M. Bishop, H. E. Varmus, unpublished data. Rabbit antisera were raised against gel-purified RSV p19²⁴.
33. G. Ramsay, G. I. Evan, J. M. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* 81, 7742 (1984).
34. M. Muller and G. Blobel, *ibid.*, p. 421.
35. We thank Irene Fecyc and Gunter Blobel for supplying the *E. coli* S135 extract and Pablo Garcia and Peter Walter for providing the *lpp* clone. We thank John Atkins, Bob Weiss, and Ray Gesteland for helpful discussions and materials and Janine Marinos for assistance in preparing the manuscript. Supported by a UCSF Earle C. Anthony/Dean's Fellowship (T.E.J.), American Cancer Society Research Professorship (H.E.V.), and grants from the National Institutes of Health and the American Chemical Society.

28 August 1985; accepted 30 October 1985

RESEARCH ARTICLE

Detection of Single Base Substitutions by Ribonuclease Cleavage at Mismatches in RNA:DNA Duplexes

Richard M. Myers, Zoia Larin, Tom Maniatis

Physical methods for detecting single base substitutions have provided powerful tools for the analysis of human genetic diseases (1-4) and the establishment of human genetic linkage maps (5-7). These techniques could also be of considerable value in the detection and analysis of single base mutations in regulatory or protein-coding sequences. Procedures available for detecting base substitutions rely on differences in restriction endonuclease cleavage sites (8-12), or on differences in the melting behavior of wild-type and mutant DNA duplexes (13-21). For example, some single base substitutions result in the loss or gain of a restriction endonuclease cleavage site, and can therefore be detected in Southern blotting experiments (8-12). However,

it is usually necessary to use a large number of different restriction enzymes before a change is detected. In addition, many substitutions cannot be detected by this procedure because they do not alter a restriction site. Another approach involves the use of synthetic oligodeoxyribonucleotides as differential hybridization probes (13-16). In this method, a labeled synthetic oligonucleotide homologous to the mutant or wild-type DNA is hybridized to blotted genomic DNA. Hybridization or washing conditions are then adjusted to allow the differential melting of the mismatched and perfectly paired duplexes. This method is useful for scoring substitutions at specific locations, but is not practical for screening large regions of DNA for new mutations or polymorphisms.

Differential DNA melting is also the basis for detecting single base substitutions by denaturing gradient gel electrophoresis (17-21). In this method, wild-type and mutant DNA molecules are separated by electrophoresis in poly-

acrylamide gels containing a gradient of formamide and urea. Duplex DNA fragments move through these gels with a constant mobility determined by molecular weight until they migrate into a portion of the gel containing a denaturant concentration sufficient to melt the DNA. When the DNA undergoes melting, its electrophoretic mobility abruptly decreases. Thus, the final position of a DNA fragment in the gel is determined by its melting temperature. The difference in melting temperature between two fragments that differ by a single base change is sufficient to allow separation on the gel. Even greater separation is achieved with DNA duplexes containing a single base mismatch (18). With specially designed plasmid vectors, virtually all possible single base substitutions can be detected in cloned DNA fragments (19, 20). However, for technical reasons (18-21), only 25 to 40 percent of all possible substitutions can be detected directly in total genomic DNA.

Because of the limitations in the procedures discussed above, we developed an alternative method for detecting single base substitutions in cloned and genomic DNA. This method involves the enzymatic cleavage of RNA at a single base mismatch in an RNA:DNA hybrid. The strategy used is based on the development of methods for synthesizing RNA probes (22-24), and on the observation that many ribonucleases are specific for single-stranded RNA under appropriate reaction conditions (25). A similar strategy had been developed earlier to detect mutations in duplex DNA containing single base mismatches (26,

Richard M. Myers is a postdoctoral fellow, Zoia Larin is a research assistant, and Tom Maniatis is a professor in the Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138. After 1 January the address for R.M.M. is Department of Physiology, University of California, San Francisco 94143.

27). In this case, attempts were made to cleave DNA:DNA mismatches with the single-strand specific nuclease S1. However, only a small amount of cleavage occurs at a few mismatches while most mismatches are not cleaved at all (26, 27). In this article, we demonstrate that many single base mismatches in RNA:DNA hybrids are cleaved specifically by ribonuclease A (RNase A).

The steps in the RNase cleavage procedure are outlined in Fig. 1. A ^{32}P -labeled RNA probe is synthesized from a wild-type DNA template with the SP6 transcription system (22–24, 28). The RNA probe is hybridized to denatured test DNA (29) in solution, and the resulting RNA:DNA hybrid is treated with RNase A (30). The RNA products are then analyzed by electrophoresis in a denaturing gel. If the test DNA is identical to wild-type DNA, a single band is observed in the autoradiogram of the gel, since the RNA:DNA hybrid is not cleaved by RNase. However, if the test DNA contains a single base substitution that results in a mismatch recognized by RNase A, two new RNA fragments will be detected. The total size of these fragments should equal the size of the single RNA fragment observed with wild-type DNA. Thus, the mutation can be localized relative to the ends of the RNA probe by determining the sizes of the cleavage products. The end of the RNA probe mapping nearest to the substitution can be determined when the experiment is performed with DNA digested by an additional restriction enzyme (29), thus localizing the substitution unambiguously.

For convenience, single base mismatches in the RNA:DNA hybrids are presented as X:Y, where X and Y designate the mismatched RNA and DNA bases, respectively. For example, "C:A" refers to a mismatch in which cytosine appears in the RNA strand opposite adenine in the DNA strand.

Detection of single base substitutions in cloned DNA fragments. To establish optimal conditions for recognizing single base mismatches, and to determine which types of mismatches can be cleaved by RNase, we examined a large number of single base substitutions in the mouse β -major globin promoter region (21, 31). With this collection, it was possible to examine all 12 types of mismatches possible in RNA:DNA hybrids in several different sequence contexts. The results of several RNase cleavage reactions are shown in Fig. 2. The RNA probe used in these reactions is complementary to the sense strand of the β -globin promoter, and therefore is designed

Abstract. Single base substitutions can be detected and localized by a simple and rapid method that involves ribonuclease cleavage of single base mismatches in RNA:DNA heteroduplexes. A ^{32}P -labeled RNA probe complementary to wild-type DNA is synthesized *in vitro* and annealed to a test DNA containing a single base substitution. The resulting single base mismatch is cleaved by ribonuclease A, and the location of the mismatch is then determined by analyzing the sizes of the cleavage products by gel electrophoresis. Analysis of every type of mismatch in many different sequence contexts indicates that more than 50 percent of all single base substitutions can be detected. The feasibility of this method for localizing base substitutions directly in genomic DNA samples is demonstrated by the detection of single base mutations in DNA obtained from individuals with β -thalassemia, a genetic disorder in β -globin gene expression.

an "antisense" probe. When this probe was annealed to the wild-type promoter fragment and then digested with RNase A, a single, full-length RNA fragment of 186 nucleotides (nt) was observed (Fig. 2, lane 1). In some experiments, faint background bands were visible in the wild-type lane, indicating that a low level of cleavage occurs at bases that are not mismatched. In contrast, when an RNA:DNA duplex containing a C:A mismatch at position -40 in the promoter was analyzed, three bands were observed (Fig. 2, lane 2). One of these bands, representing about 50 percent of the total radioactivity in the lane, corresponds to the full-length RNA probe. The lengths of the other two RNA fragments correspond to the sizes expected for cleavage at the mismatch at position -40 in the promoter (66 and 120 nt). In this and other mismatches examined, one of the RNA fragments (the 66-nt

fragment) appears as a doublet on the autoradiogram, which is probably the result of further reaction of RNase at pyrimidines near the ends of the cleaved RNA product.

Similar results were obtained with another C:A mismatch located at position -60 in the promoter (Fig. 2, lane 4). In contrast, in the case of a third C:A mismatch, occurring at -56 in the promoter, all of the radioactivity is present in the two cleavage products, indicating that 100 percent of the mismatches were cleaved under the same conditions (Fig. 2, lane 3). Altogether, 21 different C:A mismatches in the promoter were tested, and more than 50 percent of each mismatch was cleaved by RNase A in every case (Table 1). Similar results were obtained with C:C and C:T mismatches (Fig. 2, lanes 5 to 7, and Table 1). In contrast, only six of ten U:G mismatches in the promoter were cleaved by RNase, and the efficiency of cleavage varied from 10 to 90 percent (Fig. 2, lanes 8, 9, and 11, and Table 1). Three U:C mismatches were tested, and in each case cleavage was very inefficient (only 5 to 10 percent; lane 10 and Table 1). Three U:T mismatches in the promoter were cleaved at a level of 25 percent (Table 1).

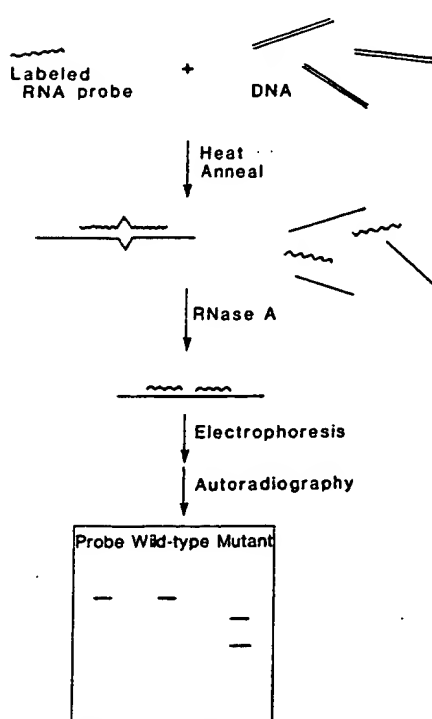


Fig. 1. Detection and localization of single base substitutions by the RNase cleavage procedure. A labeled RNA probe is synthesized with the use of the SP6 transcription system. Double-stranded DNA is digested with restriction enzymes that cleave outside the region covered by the probe and then denatured and annealed to a large molar excess of the RNA probe. Digestion of the hybridization mixture with RNase removes all of the unhybridized RNA probe and cleaves the specific RNA:DNA duplex at the position of the mismatched base. The RNase resistant products are then size-fractionated by gel electrophoresis and detected by autoradiography. In the absence of a mutation, the full-length probe fragment is observed. If the test DNA contains a single base mutation, cleavage at the resulting mismatch generates two RNA fragments whose total lengths are equal to that of the probe.

Several G:G, G:A, G:T, A:A, A:C, and A:G mismatches were tested and no cleavage by RNase A was observed in most cases (for example, see Fig. 2, lane 15). However, a small amount (10 to 20 percent) of cleavage occurred at two A:A mismatches and one G:T mismatch, and efficient cleavage occurred at three A:C and two A:G mismatches (Table 1). It is surprising that cleavage occurred at these mismatches since RNase A cleaves after pyrimidines (25). However, it is possible that destabilization of the mismatched RNA:DNA duplex leads to cleavage at nearby pyrimidine bases.

To determine whether this procedure can be used to detect small deletions, we analyzed several promoter fragments containing different single base deletions. In each case, nearly complete cleavage of the probe was observed at the resulting single base "loop-out," or at nearby pyrimidines (Fig. 2, lane 12, and Table 1). Similarly, RNA:DNA duplexes containing two mismatches in close proximity were efficiently cleaved in the assay (Fig. 2, lane 13, and Table 1).

Detection of β -thalassemia mutations in

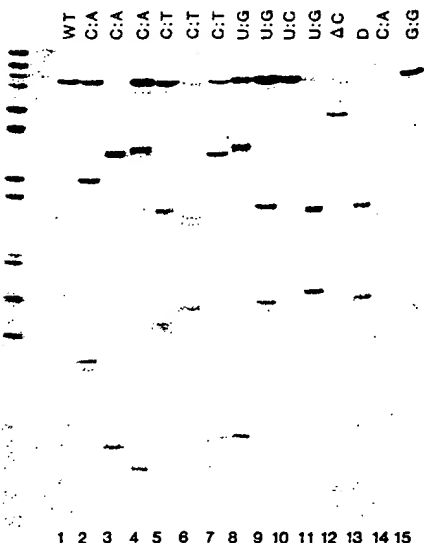


Fig. 2. RNase cleavage analysis of single base mutations in a cloned mouse β -major globin promoter fragment. A 186-nt antisense RNA probe was annealed to wild-type and mutant promoter fragments and the resulting RNA:DNA duplexes treated with RNase A. The digestion products were analyzed by polyacrylamide gel electrophoresis and autoradiography. The DNA sample analyzed in each lane was: wild-type (lane 1); mutant -40A (lane 2); -56A (lane 3); -60A (lane 4); -33T (lane 5); -25T (lane 6); -54T (lane 7); -57G (lane 8); -31G (lane 9); -62C (lane 10); -26G (lane 11); -76 deletion (lane 12); -28G/-26G (lane 13); -25A (lane 14); -49G (lane 15). The type of mismatch produced by annealing the wild-type antisense RNA probe to each mutant DNA fragment is indicated at the top of each lane.

cloned and genomic DNA. To establish the feasibility of detecting single base mutations associated with human genetic diseases, we analyzed a number of different cloned and genomic DNA's bearing β -thalassemia or sickle cell anemia mutations. In these experiments, the RNA probes used were about 615 nt in length, spanning the region of the gene and 5' flanking sequences from -128 to +485 (32). Two RNA probes were synthesized to test both the sense and antisense strand of the region. With this set of substitutions and probes, 10 of the 12 types of RNA:DNA mismatches could be formed, and 7 out of the 10 types were cleaved to some extent by RNase (Table 1).

To determine whether the RNase cleavage procedure could be used to detect single base substitutions in total genomic DNA, we analyzed DNA samples from two individuals with β -thalassemia. One individual carried a C to T transition at codon 39 of the β -globin gene in both chromosomes. The second individual was homozygous for the hemoglobin β^E (HbE) allele, which contains a G to A transition at codon 26 in the gene. The codon 39 (β^{39}) DNA was tested with the sense strand RNA probe, whereas the HbE DNA was tested with the antisense RNA probe. Both of these hybrids result in C:A mismatches with their corresponding probes. When a control experiment was performed with the sense probe and genomic DNA from an individual with wild-type β -globin genes, a single band appearing at the full-length position resulted (Fig. 3A, lane 1). When DNA from the individual homozygous for the β^{39} mutation was analyzed, RNA fragments 430 and 185 nt in length were observed (Fig. 3A, lane 2), indicating that cleavage at the C:A mismatch occurred at a high efficiency. Similar results were obtained with the analogous cloned DNA samples (Fig. 3A, lanes 3 and 4). In another experiment with the antisense RNA probe, genomic DNA from an individual with normal β -globin genes also resulted in a single band appearing at the full-length probe position (Fig. 3B, lane 1). Genomic DNA from a patient homozygous for the HbE allele resulted in two RNA fragments of the expected sizes of 355 and 260 nt (Fig. 3B, lane 2), again indicating complete cleavage of the mismatch by RNase A. These results were obtained with 3 μ g of total genomic DNA, and RNA probes with an [α - 32 P]GTP specific activity of 400 Ci/mmol. A signal could be clearly detected after a 24-hour autoradiographic exposure. These experiments therefore establish the feasibility of detecting single

base mutations and linked polymorphisms in genomic DNA with this method, at a level of sensitivity at least comparable with existing techniques.

Analysis of mismatch recognition. We find that 4 (C:A, C:C, C:T, and U:T) out of the 12 possible types of mismatches are recognized efficiently by RNase A in all sequence contexts tested. Thus, approximately one-third of all possible single base substitutions can be detected with an RNA probe homologous to one strand of the test DNA. This number can be doubled with the use of a second RNA probe, homologous to the opposite strand of the test DNA. For example, a

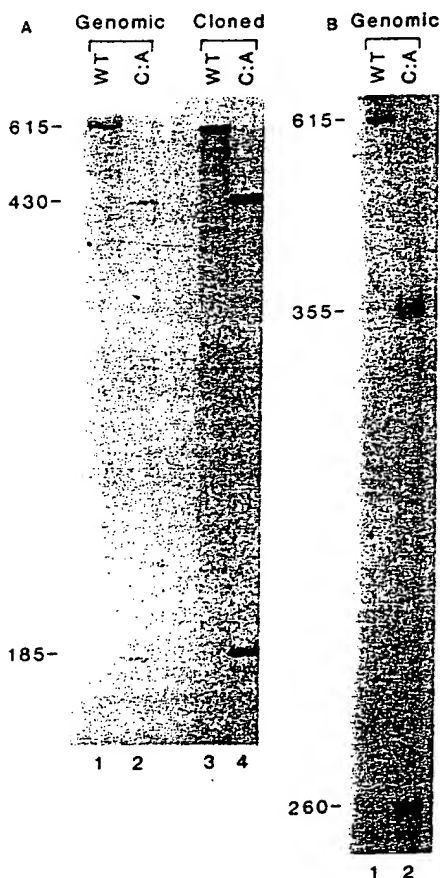


Fig. 3. RNase cleavage analysis of human genomic DNA samples from individuals with β -thalassemia. (A) Analysis of the β^{39} thalassemia mutation. Genomic and cloned DNA from an individual with wild-type β -globin genes and an individual homozygous for a nonsense mutation in codon 39 were analyzed by the RNase cleavage procedure with a sense strand RNA probe. An autoradiogram of the RNase digestion products is shown. Genomic wild-type β -globin DNA (lane 1); genomic β^{39} DNA (lane 2); cloned wild-type β -globin DNA (lane 3); cloned β^{39} DNA (lane 4). The sizes of the RNase digestion products are indicated. (B) Analysis of the HbE thalassemia mutation. Genomic DNA from an individual with wild-type β -globin genes and an individual homozygous for HbE were analyzed by the RNase cleavage procedure an antisense RNA probe. Genomic wild-type β -globin DNA (lane 1); genomic HbE DNA (lane 2).

G:T mismatch formed between one strand of the test DNA and the homologous RNA probe may not be cleaved by RNase A. However, when the other DNA strand is hybridized to its homologous RNA probe, the C:A mismatch at that same position will be cleaved by RNase. Thus, approximately two-thirds of all possible single base substitutions should be detected. This is clearly a minimum estimate, since we have observed cleavage at seven of the remaining eight possible types of mismatches in some sequence contexts.

We do not understand why some mismatches can be cleaved in some sequence contexts but not in others. It seems likely that differences in accessibility to cleavage are the result of differences in the overall structure of the mismatched duplex. However, we have not been able to discern a sequence pattern surrounding a mismatch that can be correlated with the observed efficiency of RNase cleavage.

As indicated in Fig. 2 and Table 1, some mismatches are only partially cleaved in the assay. Our data were obtained by performing the RNase reactions for a fixed length of time (30 minutes). In a time-course experiment, we found that many mismatches that are only partially cleaved in 30 minutes can be cleaved almost to completion in 90 minutes under the same conditions and with only a slight increase in background. However, mismatches not cleaved in 30 minutes are also not affected by longer incubation times. Thus, it may be desirable to perform the RNase reactions for various lengths of time in cases where partial cleavage occurs.

The temperature and ionic strength of the solution in which the RNase reaction is performed also contribute to the degree of cleavage and the apparent effects of sequence context. In fact, altering the reaction conditions to higher temperature and lower ionic strength results in cleavage at some mismatches that are not normally cleaved, and more complete cleavage of mismatches that are normally only partially cleaved. These reaction conditions may be desirable in some cases, but are not ideal since internal cleavage at perfectly matched positions also increases significantly.

The fact that some mismatches are never or rarely cleaved by RNase A and that partial cleavage sometimes occurred led us to test the ability of other ribonucleases to cleave at mismatches. We have not detected any cleavage with RNase T1 and RNase T2 under various reaction conditions.

The lack of complete cleavage of some

mismatches may pose a difficulty when the RNase cleavage procedure is used for determining the genotype of a diploid genome. In cases where 50 percent or less of the RNA probe is cleaved, low efficiency of cleavage could be an intrinsic property of the mismatch in question, or the individual may be heterozygous for the mutant allele. This ambiguity may often be eliminated by performing a time-course of RNase treatment. Alternatively, as with oligonucleotide probes (3, 16, 33), the genotype can be unambiguously determined if probes are available for both wild-type and mutant alleles. Thus, it should be possible to use

this method for prenatal diagnosis of genetic diseases. Partial cleavage at mismatches is not a problem when mapping mutations in cloned DNA samples, genomic DNA from haploid organisms, or genomic DNA sequences within the X chromosome of human males.

We have learned that a similar approach was independently developed to detect single base substitutions in messenger RNA (34). In that case ³²P-labeled antisense SP6-RNA was annealed to total cellular RNA to generate an RNA:RNA duplex containing a single base mismatch. As in the case of the RNA:DNA mismatches analyzed here,

Table 1. Tabulation of the results of an analysis of single base substitutions in the mouse β -major globin promoter region and in the human β -globin gene.

Mis-match*	Cleaved (%)†	Mutant‡	Probe§	Context	Mis-match*	Cleaved (%)†	Mutant‡	Probe§	Context
C:A	50	M -40A	AS	CCUG	G:A	0	M -49A	AS	CUGCC
	100	M -56A	AS	CUAUA		0	M -42A	AS	CUGGC
	50	M -51A	AS	GCACU		0	M -54T	S	CAGAG
	100	M -22A	AS	CUAC		0	M -35T	S	CAGAG
	100	M -19A	AS	UAACU		0	M -13T	S	GAGCA
	100	M +19A	AS	AACUA		0	M -25T	S	AAGGU
	50	M -33A	AS	UGCUC		75	H IVS15T	S	UGGUA
	50	M -50A	AS	UGCCC					
	100	M -52A	AS	CCGUC	G:T	15	M -49T	AS	CUGCC
	100	M -65T	S	CACAC		0	M -42T	AS	CUGGC
	100	M -63T	S	CACAG		0	M -50A	S	GGGCA
	75	M -49T	S	GGCAG		0	M -51A	S	AGGGC
	100	M -42T	S	GCCAG		0	M -40A	S	CAGGG
	50	M -37T	S	GGCAG		0	M -25A	S	AAGGU
	75	M -32T	S	AGCAU		0	M -22A	S	GUGAG
	50	M -4T	S	CUCCU		0	M -19A	S	AGGUA
	100	M -1T	S	CUAC		0	M -61A	S	CAGGA
	100	M +2T	S	CACAU		0	M -60A	S	AGGAU
	60	M -77T	S	GGCCA		0	M -56A	S	UAGAG
	100	H HbE	AS	CUAC		0	H Hb39	AS	CUGGG
	50	H IVS11	AS	AACCU		0	H IVS11	S	AGGUU
	100	H C39	S	CCGAG		0	H HbE	S	GUGAG
C:C	100	M -42G	S	GCCAG	G:G	0	M -49G	AS	CUGCC
	100	M -32G	S	AGCAU		0	M -42G	AS	CUGGC
	100	M -76G	S	GCCAA		0	M -76G	AS	UUGGC
	100	H IVS15C	AS	UAACU		0	M -77G	AS	UGGCC
C:T	50	M -33T	AS	GACCA		0	M +13C	S	CUAC
	75	M -25T	AS	AACGU		0	H IVS15C	S	UGGUA
	75	M -54T	AS	GACAG	A:C	0	M -23C	AS	UCACC
	100	M -65A	S	CACAC		0	M -38C	AS	CUAUC
	100	M -63A	S	CACAG		75	M -57G	S	AUAGA
	50	M -43A	S	AGCCA		0	M -48G	S	GCAGG
	100	M -32A	S	AGCAU		0	M -74G	S	CAAUC
	100	M -67A	S	CUAC		0	M -55G	S	AGAGA
	90	H IVS15T	AS	UAACU		0	M -45G	S	GGAGC
U:G	50	M -57G	AS	UCUAU		0	M -34G	S	ACAGC
	30	M -31G	AS	UAUGC		100	M -31G	S	GCUAU
	90	M -26G	AS	CCUUA		100	M -29G	S	AUAGA
	0	M -48G	AS	CCUGC		0	M +14G	S	UGACA
	0	M -45G	AS	GCUCC		50	H IVS16	AS	AUACC
	40	M -59G	AS	UAUCC	A:A	10	M -13A	AS	UGAUC
	30	M -34G	AS	CGUCU		0	M -2A	AS	UGAGG
	0	M -23C	S	GGUGA		0	M -73A	AS	UGAUU
	10	M -18C	S	GGUAG		0	M -30A	AS	UAAGU
	0	M +12C	S	UCUGA		20	M -48T	S	GCAGG
	25	H IVS16	S	GGUAU		0	H HbS	S	UGAGG
U:C	5	M -62C	AS	CCUGU	A:G	100	M -66C	S	UCACA
	5	M -66C	AS	UGUGA		50	M -62C	S	UCAGG
	5	M -45C	AS	GCUCC		0	M -45C	S	GGAGC
U:T	25	M -13A	S	GAUCA	Deletion	100	M +2CD	AS	AUGUG
	25	M -2A	S	CCUCA		100	M -76CD	AS	UUGGC
	25	M +6A	S	UUUGC		100	M +10TD	AS	CUUCU
	75	H HbS	AS	CCUCU		100	M -18TD	AS	GGUAG
					Double	100	-28G/-26G	AS	
						100	-28C/-23C	AS	

*The type of mismatch formed in each case. †The fraction of the total protected RNA probe that is present in cleaved RNA fragments. ‡The mouse promoter mutants are indicated by M followed by a number designating their position relative to the cap site of β -globin transcription (32). The human β -thalassaemia mutations are indicated by H followed by the mutation. §The probe used in each case is designated either as sense (S) or antisense (AS). ||The nucleotides surrounding each mismatch in the RNA strand are indicated in a 5' to 3' direction. The underlined nucleotide in each case occurs at the position of the mismatch.

RNA:RNA mismatches are also cleaved by RNase A.

Applications. The RNase cleavage procedure described provides a sensitive, rapid, and simple means of detecting single base substitutions in cloned or genomic DNA. The ^{32}P -labeled RNA probes are easily prepared with well-characterized SP6-plasmid vectors, the required enzymes are commercially available, and the electrophoresis involves the use of standard DNA sequencing gels. In addition, analysis of the sizes of the RNase cleavage products of the RNA:DNA heteroduplexes not only provides evidence for the presence of a single base mismatch in the test DNA but also makes it possible to localize the mismatch to within a few nucleotides.

The RNase cleavage procedure should be applicable to problems where the detection and localization of single base substitutions is important. For example, the procedure can be applied to the analysis of human genetic diseases. By establishing sets of SP6-plasmids containing DNA fragments that span an entire gene, it should be possible to survey rapidly even the largest genes for single base mutations. Similarly, this method should be valuable for detecting neutral polymorphisms in genetic linkage studies. The ability to detect a large fraction of all possible single base substitutions in a DNA fragment with a single RNA probe represents a significant advance over current methods that involve the detection of restriction fragment length polymorphisms. Another application of this procedure is the localization of mutations that are genetically selected.

References and Notes

1. S. H. Orkin and H. H. Kazazian, *Annu. Rev. Genet.* 18, 131 (1984).
2. M. A. Lehrman, J. L. Goldstein, M. S. Brown, D. W. Russell, W. J. Schneider, *Cell* 41, 735 (1985).
3. V. J. Kidd, R. B. Wallace, K. Itakura, S. L. C. Woo, *Nature (London)* 304, 230 (1983).
4. J. Gitschier, D. Drayna, E. G. D. Tuddenham, R. L. White, R. M. Lawn, *ibid.* 314, 738 (1985).
5. E. Solomon and W. F. Bodmer, *Lancet* 1979-I, 923 (1979).
6. D. Botstein, R. White, M. Skolnick, R. Davis, *Am. J. Hum. Genet.* 32, 314 (1980).
7. R. L. White *et al.*, *Nature (London)* 313, 101 (1985).
8. R. A. Flavell, J. M. Kooter, E. DeBoer, P. F. R. Little, R. Williamson, *Cell* 15, 25 (1978).
9. J. C. Chang and Y. W. Kan, *Lancet* 1981-II, 1127 (1981).
10. R. F. Geever, L. B. Wilson, F. S. Nallaseth, P. F. Milner, M. Bittner, J. T. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* 78, 5081 (1981).
11. J. C. Chang and Y. W. Kan, *N. Eng. J. Med.* 307, 30 (1982).
12. S. H. Orkin, P. F. R. Little, H. H. Kazazian, C. D. Boehm, *ibid.*, p. 32.
13. R. B. Wallace *et al.*, *Nucleic Acids Res.* 6, 3543 (1979).
14. R. B. Wallace *et al.*, *ibid.* 9, 879 (1981).
15. R. B. Wallace, M. J. Johnson, P. Dembek, K. Itakura, *ibid.*, p. 3647.
16. M. Pirastu *et al.*, *N. Eng. J. Med.* 309, 284 (1983).
17. S. G. Fischer and L. S. Lerman, *Proc. Natl. Acad. Sci. U.S.A.* 80, 1579 (1983).
18. R. M. Myers, N. Lumelsky, L. S. Lerman, T. Maniatis, *Nature (London)* 313, 495 (1985); R. M. Myers, T. Maniatis, L. S. Lerman, in preparation.
19. R. M. Myers, S. G. Fischer, T. Maniatis, L. S. Lerman, *Nucleic Acids Res.* 13, 3111 (1985).
20. R. M. Myers, S. G. Fischer, L. S. Lerman, T. Maniatis, *ibid.*, p. 3131.
21. R. M. Myers, L. S. Lerman, T. Maniatis, *Science* 229, 242 (1985).
22. M. R. Green, T. Maniatis, D. A. Melton, *Cell* 32, 681 (1983).
23. K. Zinn, D. DiMaio, T. Maniatis, *ibid.* 34, 865 (1983).
24. D. A. Melton, P. A. Krieg, M. R. Rebagliati, T. Maniatis, M. R. Green, *Nucleic Acids Res.* 12, 7035 (1984).
25. G. G. Brownlee, in *Laboratory Techniques in Biochemistry and Molecular Biology*, T. S. Work and E. Work, Eds. (American Elsevier, New York, 1972), pp. 67-99.
26. T. E. Shenk, C. Rhodes, P. W. J. Rigby, P. Berg, *Proc. Natl. Acad. Sci. U.S.A.* 72, 989 (1975).
27. J. B. Dodgson and R. D. Wells, *Biochemistry* 16, 2374 (1977); B. Seed, personal communication; R. M. Myers, unpublished data.
28. Uniformly labeled single-stranded RNA probes were made as described (24) with [^{32}P]GTP as the only labeled nucleotide. Bacteriophage SP6 RNA polymerase and [^{32}P]GTP were purchased from New England Nuclear. RNasin was from Promega Biotec and nucleoside triphosphates were from PL Biochemicals. Probes used with cloned DNA samples were synthesized with [^{32}P]GTP (40 Ci/mmol, 100 μM) in the transcription reaction, and could be used for a period of 2 weeks with little evidence of degradation. Probes used with genomic DNA samples contained [^{32}P]GTP (400 Ci/mmol) and were used within 3 days. After synthesis, the probe made from 1 μg of template DNA was resuspended in 150 μl of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 0.1 percent SDS. Approximately 0.5×10^5 cpm of probe, which contained 5×10^4 to 1×10^5 cpm or 5×10^5 to 1×10^6 cpm for probes made for cloned or genomic DNA samples, respectively, was used in each hybridization reaction. RNA probes used in this procedure must be full-length to obtain unambiguous cleavage results. With some DNA sequences it is difficult to achieve 100 percent full-length probe synthesis in the runoff transcription reaction, particularly when synthesizing genomic probes where the GTP concentration is low. In these cases, the full-length RNA probe should be purified by gel electrophoresis after synthesis.
29. To prepare a cloned or genomic DNA sample for RNase cleavage analysis, the DNA was digested with restriction enzymes that do not cleave within the sequence homologous to the RNA probe. By performing the annealing and cleavage reactions with this DNA and analyzing the RNA products by gel electrophoresis, the position of a substitution can be mapped accurately relative to one or the other of the two ends of the probe. To determine unambiguously which end of the probe is nearest to the substitution, a second experiment can be performed as follows. The DNA sample is digested with a restriction enzyme that cleaves once within the sequence homologous to the probe, generating an additional DNA fragment. After the hybridization reaction, each RNA:DNA duplex will contain a single-stranded "overhang" of RNA probe that will be digested to oligonucleotides in the RNase reaction. Analysis of the RNA products by electrophoresis will result in the replacement of one of the RNA's with two species whose total size equals that of the missing fragments, revealing the exact position of the substitution relative to the restriction site. The hybridization is carried out with the RNA probe in molar excess to avoid rejoining the two DNA fragments released by restriction digestion by hybridization to the same RNA molecule.
30. Ribonuclease A (Sigma, R-5125) was dissolved in distilled water to a concentration of 2 mg/ml and then placed in a boiling water bath for 10 minutes. After cooling to room temperature, the solution could be stored at $+4^\circ\text{C}$ for as long as 1 year. Formamide (MCB, Inc.) was deionized by stirring with Dowex Mixed Bed Resin AG501XB (Bio-Rad) at room temperature for 30 minutes and was stored at $+4^\circ\text{C}$. Hybridizations were performed by suspending 20 to 50 ng of cloned plasmid DNA or 3 to 6 μg of total genomic DNA in 30 μl of hybridization buffer [80 percent formamide, 40 mM Pipes (pH 6.4), 0.4M NaCl, and 1 mM EDTA], adding 0.5 μl of labeled RNA probe (28) and treating the mixture at 90°C for 10 minutes. The samples were then incubated at 45°C for 30 minutes (cloned DNA) or 10 hours (genomic DNA). After the annealing, 350 μl of a solution containing RNase A (40 $\mu\text{g}/\text{ml}$) in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 200 mM NaCl, and 100 mM LiCl was added to the hybridization mixture. The sample was then mixed by vortexing and incubated at 25°C for 30 minutes. The RNase reaction was stopped by the addition of 10 μl of 20 percent SDS and 10 μl of proteinase K (10 mg/ml) (Boehringer Mannheim) followed by incubation at 37°C for 15 to 30 minutes. Protein was then removed by extraction with an equal volume of phenol-chloroform (1:1) containing 4 percent isoamyl alcohol and 0.01 percent hydroxyquinoline. The aqueous supernatant (300 μl) was removed without disturbing the interface to ensure that no traces of RNase remained. After the addition of 10 to 20 μg of carrier transfer RNA, the samples were precipitated with ethanol and resuspended in 5 μl (genomic DNA) or 25 μl (cloned DNA) of loading buffer and analyzed by denaturing polyacrylamide gel electrophoresis [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), p. 184]. A time-course of annealing the RNA probe to genomic DNA indicated that the hybridization was complete after 6 hours with the quantities of probe and genomic DNA (3 to 6 μg) used here. Because there may be variations in probe and genomic DNA concentration in different experiments, 10-hour hybridizations were routinely performed to ensure maximum signals. Reaction conditions for RNase cleavage were chosen empirically. We found that a buffer containing 100 mM LiCl and 200 mM NaCl suppressed the background to a greater extent than did 300 mM NaCl alone. In addition, a slightly higher temperature (30°C) of RNase treatment resulted in more complete cleavage at partially cleaved mismatches, but also caused some increase in background. Results very similar to our previous result with RNase treatment at 16°C in a buffer containing 100 mM LiCl and 100 mM NaCl.
31. D. A. Konkel, S. M. Tilghman, P. Leder, *Cell* 15, 1125 (1978). The RNA probes used in the analysis of the mouse β -globin promoter were derived from SP6 plasmids containing a 186-bp Hinf I-Dde I fragment (from -106 to +72 relative to the mRNA cap site) of the mouse β -major globin gene. The sense and antisense probes were obtained by inserting this fragment into pSP64 and pSP65 in both orientations relative to the bacteriophage promoter. The promoter insert contains eight extra base pairs due to the addition of a Bgl II linker at position +26 in the gene. This additional 8-bp linker is also present in the plasmids carrying the mutations. The collection of single base substitutions in the mouse β -major globin promoter region was generated by a random chemical method followed by purification by denaturing gradient gel electrophoresis (21).
32. R. M. Lawn, E. F. Fritsch, R. C. Parker, G. Blake, T. Maniatis, *Cell* 15, 1157 (1978). The RNA probes used in the analysis of the human β -thalassaemia mutations were derived from plasmids containing a 605 bp Rsa I to Bam HI fragment (from positions -128 to +477 relative to the mRNA cap site) of the human β -globin gene. The sense and antisense probes were obtained by inserting this fragment into pSP64 and pSP65 in both orientations relative to the bacteriophage promoter. The plasmids carrying cloned human β -globin genes containing the normal allele or the thalassaemia mutations have been described (1). A plasmid carrying the sickle allele was made from a bacteriophage λ clone provided by S. Orkin. Genomic DNA samples from β -thalassaemia patients were provided by H. Kazazian.
33. S. H. Orkin, A. F. Markham, H. H. Kazazian, *J. Clin. Invest.* 71, 775 (1983).
34. E. Winter, E. Yamamoto, C. Almoguera, M. Perucho, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
35. We thank H. Kazazian for providing us with genomic DNA samples of defined genotype; S. Goodbourn and B. Seed for helpful discussions and suggestions for reaction conditions; P. Krieg and A. Krainer for comments on the manuscript, and J. Ma and M. Green for contributing to the early stages of this work. Supported by grants from the National Institutes of Health (T.M.), The Wills Foundation, and a Special Fellowship of the Leukemia Society of America (R.M.M.).

6 September 1985; accepted 28 October 1985

Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)

C.R.Newton, A.Graham, L.E.Heptinstall, S.J.Powell, C.Summers, N.Kalsheker¹, J.C.Smith and A.F.Markham

ICI Diagnostics, Gadbrook Park, Northwich, Cheshire CW9 7RA. ICI Cellmark Diagnostics, Blacklands Way, Abingdon Business Park, Abingdon, Oxon OX14 1DY, ¹University of Wales College of Medicine, Department of Medical Biochemistry, Newport Road, Cardiff CF2 1SZ. UK and ICI Cellmark Diagnostics, 20271 Golden Rod Lane, Germantown, MD 20874, USA

Received February 23, 1989; Accepted March 2, 1989

ABSTRACT

We have improved the "polymerase chain reaction" (PCR) to permit rapid analysis of any known mutation in genomic DNA. We demonstrate a system, ARMS (Amplification Refractory Mutation System), that allows genotyping solely by inspection of reaction mixtures after agarose gel electrophoresis. The system is simple, reliable and non-isotopic. It will clearly distinguish heterozygotes at a locus from homozygotes for either allele. The system requires neither restriction enzyme digestion, allele-specific oligonucleotides as conventionally applied, nor the sequence analysis of PCR products. The basis of the invention is that unexpectedly, oligonucleotides with a mismatched 3'-residue will not function as primers in the PCR under appropriate conditions. We have analysed DNA from patients with α -antitrypsin (AAT) deficiency, from carriers of the disease and from normal individuals. Our findings are in complete agreement with allele assignments derived by direct sequencing of PCR products.

INTRODUCTION

The analysis of nucleic acid sequence is central to biology. Determination of variation in DNA sequence between individuals underpins molecular genetics. Such analysis is routinely performed by examination of restriction fragment length polymorphism (RFLP) using the Southern blotting technique (1,2,3). This approach has proved enormously useful, generating a massive literature, despite the fact that it is relatively slow and only allows for the examination of the limited number of polymorphic base changes which either create or destroy a restriction endonuclease recognition site. Without doubt any method which enabled all polymorphic base changes in the genome to be examined in a facile manner would be invaluable to the molecular genetics community.

PCR (4) has greatly facilitated the analysis of genomic DNA. It allows diagnosis of genetic diseases when combined with one of a variety of other techniques (5,6,7,8,9,10,11,12). We and others have reported the use of PCR and direct sequencing for diagnosis of inherited diseases (5,6,7,13). Allele-specific oligonucleotides (ASOs), (14,15) either radio-labelled (8) or non-isotopically tagged (9) have been applied to disease diagnosis in the conventional manner by probing dot blots of PCR products. Occasionally a point mutation giving rise to a specific phenotype may create or destroy a restriction enzyme recognition site (2). In such instances PCR products may (or may not) be cleaved when treated with the restriction enzyme. The presence or absence of the restriction site can be used to perform diagnoses as recently demonstrated for sickle cell anaemia (10). Similarly a polymorphic restriction site may be in linkage with an uncharacterised mutation allowing diagnoses to be performed in informative families by analysis of the amplified restriction site polymorphism (11,16,17).

We demonstrate here a general technique which allows the scrutiny of any point mutation polymorphism. The technique requires that the terminal 3'-nucleotide only of a PCR primer be allele specific. Thus the primer is synthesised in two forms. The 'normal' form is refractory to PCR on 'mutant' template DNA and the 'mutant' form is refractory to PCR on 'normal' DNA. In some instances a single 3'-mismatched base does allow amplification to proceed. We have shown that introducing additional deliberate mismatches near the 3' end of appropriate primers ameliorates this problem.

Molecular characterisation of the genes associated with the more common inherited disorders is constantly providing new information about the underlying, disease-associated mutations. Indeed recent sequencing of mutant β -globin genes has only rarely resulted in the discovery of novel alleles (18). This implies that at the β -globin locus characterisation of the molecular pathology is nearing completion (13). Diseases such as cystic fibrosis, as yet uncharacterised at the gene level, may have several RFLPs in linkage disequilibrium with the affected phenotype (19). Such RFLPs are useful for haplotype analysis and

risk assessment of carrier status particularly where there is a family history of the disease (20). Furthermore, flanking sequences of some such RFLPs have been determined allowing PCR followed by restriction analysis for haplotype identification (16,17). Some concern has been expressed as to the reliability and reproducibility of such assays in the absence of rigorous and appropriate internal controls. In theory ARMS would allow rapid haplotype analysis in such situations, providing sufficient genomic sequence is known around the polymorphic restriction site.

The feasibility of our ARMS was demonstrated by the amplification of exon III and part of intron III in the human AAT gene (figure 1). Direct application of ARMS to the clinically significant S and Z alleles of AAT (21) was performed and the diagnoses were in agreement with the results of sequence analysis of the PCR products (5).

MATERIALS AND METHODS

DNA preparation

Genomic DNA was isolated from peripheral blood cells as described previously (5).

Oligonucleotide amplification, amplification refractory and sequencing primers

The common primers 1,2,5 and 6 (figure 1) were those described previously (5). Their respective sequences were d(CCCACCTTCCCCTCTCTCCAGGCAAATGGG), d(GGGCCTCAGTCCCAACATGGCTAAGAGGTG), d(TGTCCA CGTGAGCCTTGCTCGAGGCCTGGG) and d(GAGACTTGGTATTTTGTTCATCATTAAAG). Primer 2a and the 3,4,7 and 8 series of primers (figure 2) as well as the primers for the internal control, a 510 base pair fragment from the unusually long exon 26 of the human apolipoprotein B gene (22) were prepared as described (5) and were used without further purification. The sequencing primers for initial allele characterisation were those described earlier (5).

Allele characterisation by PCR and direct sequencing

Mutant and normal alleles of the AAT S and Z loci were confirmed by PCR amplification either as described (5), or as follows; target sequences were amplified in a 100 μ l reaction volume

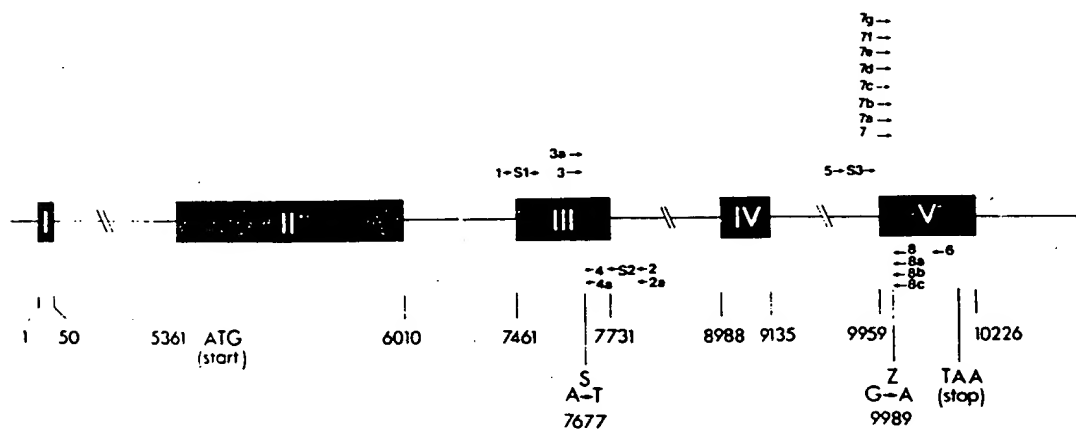


Figure 1

The human alpha-1-antitrypsin gene. Coordinates are as described by Long *et al.* (24). Position 1 is the proposed transcription start site. The solid boxes represent the five exons, the S and Z loci are shown, as are the respective mutations responsible for the S and Z phenotypes. The arrows below and above the gene represent the various primers used. Primers prefixed by S are those used in direct sequencing of PCR products to confirm genotypes prior to ARMS analyses. The remaining primers are those used to demonstrate the feasibility of the ARMS concept and those used in ARMS analyses, these primers are shown in detail in figure 2.

containing approximately 1µg genomic DNA, deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP) and thymidine triphosphate (TTP), each 1.5mM, 67mM Tris-HCl (pH8.8), 16.6mM ammonium sulphate, 6.7 mM magnesium chloride, 10mM 2-mercaptoethanol, 6.7µM EDTA and 1µM each appropriate amplification primer. Samples were heated at 100°C for 5 minutes to denature the DNA. Two units of *Thermus aquaticus* (Taq) DNA polymerase (23) (Anglian Biotechnology) were added to each sample. Samples were overlaid with light mineral oil (Sigma, 50µl) then heated at 60°C for 4 minutes for the first round of DNA synthesis. (See Discussion). Subsequent cycles consisted of a two minute denaturation step at 92°C and a combined primer annealing and DNA synthesis step at 60°C for 4 minutes. 33 cycles were performed and the DNA synthesis step of the final cycle was extended to 20 minutes. Direct sequencing of the PCR products was as described previously (5).

ARMS analysis of genomic DNA

The feasibility of the ARMS concept was demonstrated using duplicate samples of genomic DNA from one normal individual. AAT exon III primers 1 and 2 were used with one sample; primers 1 and 2a (3'C/T mismatch) were used with the other sample. AAT exon V primers 5 and 6 were present in both samples serving as an internal control. All primers are shown in figure 1. PCR reactions were performed and examined by agarose gel electrophoresis (3% Nu-sieve) as previously described (5).

In applying ARMS to subsequent mutation analysis, primers 'Control 1' d(CTCTGGGAGCACAGTACGAAAAACCACTT) and 'Control 2' d(AA TGAATTTATCAGCCAAAACCTTTTACAGG) were included in all reactions and served to provide an internal control PCR product. The control primers amplify a 510 base pair product within exon 26 of the human apolipoprotein B gene (22).

Genomic DNAs of characterised AAT genotypes MM,MS,MZ and ZZ were subjected to PCR so as to amplify the internal control fragment. In separate pairs of reactions each DNA was either coamplified with the appropriate 'normal' or 'mutant' primer paired with a common primer for the respective AAT locus. These primers are shown in figure 2.

The reactions for the ARMS analyses were performed in a volume of 100 μ l containing approximately 1 μ g genomic DNA. dATP, dCTP, dGTP and TTP were each 1.5mM in 67mM Tris-HCl (pH8.8), 16.6mM ammonium sulphate, 6.7mM magnesium chloride, 10mM 2-mercaptoethanol, 6.7 μ M EDTA and 1 μ M each appropriate amplification primer. Samples were heated at 100°C for 5 minutes to denature the DNA. Two units Taq DNA polymerase (Anglian Biotechnology) was added to each sample. Samples were overlaid with light mineral oil (Sigma, 50 μ l) then heated at 60°C for 4 minutes for the first round of DNA synthesis. Subsequent rounds of amplification comprised two minutes denaturation at 92°C followed by combined primer annealing and DNA synthesis at 60°C for four minutes. 33 cycles were performed in this way with the final synthesis step extended to 20 minutes. 18 μ l from each reaction was combined with 2 μ l of 50% glycerol 0.2% bromophenol blue in 1X TBE then electrophoresed on 1.4% agarose gels in 1X TBE containing 0.5 μ g/ml ethidium bromide.

```

5' ATTCCCAACCTGAGGGTGACCAAGAAGCTGCCACACCTCTTAGCCATGTTGGGACTGAGGCCCATCAGGACTGGC 3'
3' TAAGGGTTGGACTCCCACTGGTTCTTCGACGGGTGTGGAGAATCGGTACAACCCTGACTCCGGGTAGTCCTGACCG 5'
    |                                     Primer 2   GTGGAGAATCGGTACAACCCTGACTCCGGG 5'
7735                                     Primer 2a  TTGGAGAATCGGTACAACCCTGACTCCGGG 5'                                     7811

5' GCCTGATGAGGGGAAACTACAGCACCTGGT Primer 3a
5' GCCTGATGAGGGGAAACTACAGCACCTGGA Primer 3
5' CTTCTGCCTGATGAGGGGAAACTACAGCACCTGGaAAATGAACTACCCACGATATCATACCAAGTTCCTGGAAA 3'
3' GAAGGACGGACTACTCCCTTTGATGTCGTGGACcTTTACTTGAGTGGGTGCTATAGTAGTGGTTCAAGGACCTTT 5'
    |                                     Primer 4   TTTACTTGAGTGGGTGCTATAGTAGTGGT 5'
7642                                     Primer 4a  ATTACTTGAGTGGGTGCTATAGTAGTGGT 5'                                     7718

5' CCGTGCATAAGGCTGTGCTGACCCTCGACA Primer 7g
5' CCGTGCATAAGGCTGTGCTGACCCTCGACG Primer 7f
5' CCGTGCATAAGGCTGTGCTGACCATAGACA Primer 7e
5' CCGTGCATAAGGCTGTGCTGACCATAGACG Primer 7d
5' CCGTGCATAAGGCTGTGCTGACCATCGCCA Primer 7c
5' CCGTGCATAAGGCTGTGCTGACCATCGCG Primer 7b
5' CCGTGCATAAGGCTGTGCTGACCATCGACA Primer 7a
5' CCGTGCATAAGGCTGTGCTGACCATCGACG primer 7
5' TCCAGGCCGTGCATAAGGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTAGAGGCC 3'
3' AGGTCCGGCACGTATTCCGACACGACTGGTAGCTGcTCTTCCCTGACTTCGACGACCCCGGTACAAAAATCTCCGG 5'
    |                                     Primer 8   CTCTTCCCTGACTTCGACGACCCCGGTAC 5'
9954                                     Primer 8a  TTCTTCCCTGACTTCGACGACCCCGGTAC 5'                                     10030
                                     Primer 8b  CTCATTCCTGACTTCGACGACCCCGGTAC 5'
                                     Primer 8c  TTCATTCCTGACTTCGACGACCCCGGTAC 5'

```

Figure 2

ARMS primers. The top panel shows the primers used to test the ARMS concept. Primer 2 is complementary to the coding strand of the AAT gene. Primer 2a shows the 3'-OH mismatched T residue. Primers 2 and 2a are used in conjunction with primer 1 (figure 1). The centre panel shows the ARMS primers employed at the AAT S locus. The lower case A/T base pair is the AAT S locus and the depicted sequence is the normal sequence. The AAT S variant DNA has a T/A base pair at this position. Primers 3 and 4 correspond to 'normal' sequence, primers 3a and 4a correspond to 'mutant' sequence. The lower panel shows the ARMS primers employed at the AAT Z locus. The lower case G/C base pair is the Z locus and the normal sequence is shown. The AAT Z variant DNA has an A/T base pair at this position. Primers 7 and 8 correspond to the 'normal' sequence, primers 7a and 8a correspond to the 'mutant' sequence. Primers 7, 8, 7a and 8a have not been destabilised. The remaining primers in the 7 and 8 series are destabilised and the deliberately introduced mismatches are underlined. Primers 7b, 7d, 7f and 8b correspond to 'normal' sequence (discounting the deliberate mismatches) likewise primers 7c, 7e, 7g and 8c correspond to 'mutant' sequence, again discounting the introduced mismatches. The position numbers are as described by Long *et al.* (24).

RESULTS

ARMS primers

Figure 1 shows the ARMS primers in relation to the human AAT gene. Figure 2 shows each ARMS primer sequence in detail with respect to the gene. The position numbers are measured from the proposed transcription start site of the AAT gene (24). Discounting the variable 3' nucleotides and deliberately introduced mismatches, the 2,3,7 and 8 series primers are 59% GC 30mers and the 4 series primers are 41% GC 30mers. The common primers 1,2,5 and 6 are a 63% GC 30 mer, 60% GC 30mer, 67% GC 30mer and a 31% GC 29mer respectively.

Feasibility of the ARMS concept

We have shown previously that the AAT gene regions bounded by primers 1 and 2 and by primers 5 and 6 (figures 1 and 2) can be coamplified without affecting the efficiency of amplification of either target performed in isolation (5). We chose to introduce a 3' terminal base change into primer 2. Specifically the 3' dG residue was replaced by T to provide primer 2a. This substitution generates a template/primer C/T mismatch. When primers 1,2,5 and 6 are combined in a PCR, both the 360 bp product bounded by primers 1 and 2, and the 220 bp product bounded by primers 5 and 6 are observed (figure 3, lane 1). Substitution of primer 2 by primer 2a however blocks amplification of the 360 bp product while the internal control 220 bp product is still generated (figure 3, lane 2). This result is attributable to the lack of a 3' exonucleolytic proofreading activity of Taq DNA polymerase (23) in agreement with the observations of Tindall and Kunkel (25).

ARMS analysis of the AAT S locus

Genomic DNAs, either homozygous normal with respect to the AAT gene S allele or heterozygous S were each amplified as described in Materials and Methods. Each DNA was separately amplified using primers 2 and 3 and primers 2 and 3a. Primer 3 corresponds to the normal sequence at the S locus and primer 3a corresponds to the S variant sequence. In all reactions the internal control primers were also included.

On the normal DNA substrate, product was derived only from the internal control primers and primers 2 and 3. No product was

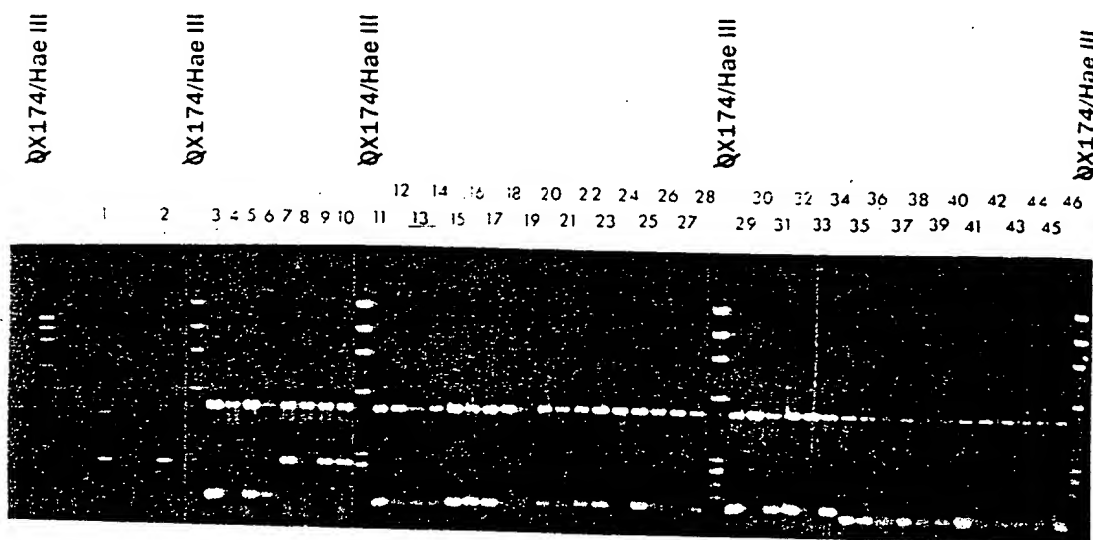


Figure 3

Agarose gels showing the feasibility of the ARMS concept (lanes 1 and 2) and ARMS analyses at the AAT S locus (lanes 3 to 10) and AAT Z locus (lanes 11 to 46). Specific reactions are as described in the text (results section).

observed when primer 3a replaced primer 3. When heterozygous S DNA replaced homozygous normal DNA the expected 152 bp product was generated when either primer 3 or 3a was included in the reaction (figure 3, lanes 3 to 6). Primer 3 generates an A/A mismatch with S variant DNA and primer 3a generates a T/T mismatch with normal DNA. When the ARMS detection primers were designed for the opposing strand at the S locus (primers 4, normal and 4a, S variant) and used for amplification with a common primer 1, similar results were obtained. The 510 bp internal control was generated but the 267 bp product was observed only when the normal primer was applied to normal DNA or the normal or S variant primer was applied to heterozygous S DNA. The 267 bp product was not generated when the S variant primer was applied to normal DNA (figure 3, lanes 7 to 10). Primer 4 generates a T/T mismatch with S variant DNA and primer 4a generates an A/A mismatch with normal DNA.

ARMS analysis of the AAT Z locus

In analogous experiments to the ARMS analyses of the AAT S locus we amplified genomic DNAs characterised as normal, heterozygous and homozygous at the AAT Z locus. All reactions contained the

internal control primers 1 and 2. Initial experiments contained primer 7 (normal) or 7a (mutant) for amplification with common primer 6 to yield a 150 bp product (figures 1 and 2). Alternative experiments targetting the opposing strand and employing primers 8 (normal) or 8a (mutant) for amplification with common primer 5 (figures 1 and 2) would give products of 129 bp. Figure 3, lanes 11-16 and lanes 35-40 shows the products generated by the respective use of primer 7 with normal DNA, primer 7a with normal DNA, primer 7 with heterozygous DNA, primer 7a with heterozygous DNA, primer 7 with homozygous Z (ZZ) DNA, primer 7a with ZZ DNA, primer 8 with normal DNA, primer 8a with normal DNA, primer 8 with heterozygous DNA, primer 8a with heterozygous DNA, primer 8 with ZZ DNA and primer 8a with ZZ DNA. In contrast to the ARMS data for the AAT S locus, the corresponding results for the AAT Z locus show reduced specificity in that products were evident using either normal primer with ZZ DNA and either mutant primer with normal DNA. Primer 7 with ZZ DNA generates a primer/template G/T mismatch. Conversely primer 7a with normal DNA generates an A/C mismatch. Primer 8 with ZZ DNA generates a C/A mismatch and primer 8a with normal DNA generates a T/G mismatch.

In an attempt to increase the specificity of the ARMS primers we chose to deliberately introduce an additional mismatch near their 3'-ends. When primers 7f and 7g which have a deliberate C/T mismatch seven bases from the 3'-end were introduced to replace primers 7 and 7a (figure 2) specificity was improved. Figure 3, lanes 17-22 shows the products of these reactions. In particular, lane 18 shows the virtual absence of the 150 bp product when the mutant primer (7g) is applied to normal DNA. Unfortunately the normal primer (7f) still generates a small amount of product with ZZ DNA (figure 3, lane 21), but much reduced with respect to the yield with the mutant primer on ZZ DNA with equivalent internal control products. When primers 7d and 7e which have a deliberate A/G mismatch five bases from their 3'-ends were introduced to replace primers 7 and 7a similar results were obtained (figure 3 lanes 23 to 28) to those with primers 7f and 7g. When primers 7b and 7c replaced primers 7 and 7a in the system the desired specificity was observed. Primers 7b and 7c have a deliberate C/T mismatch three bases from their 3'-ends. Specifically only primer

7b generated a 150 bp product with normal DNA (fig. 3 lane 29). The 'mutant' primer 7c failed to do so (fig. 3 lane 30). Both primers 7b and 7c generated product from heterozygous DNA (fig. 3 lanes 31 and 32). Primer 7b failed to generate the 150 bp product with ZZ DNA whereas the 'mutant' primer 7c did generate the 150 bp product (fig. 3 lanes 33 and 34). Similar exchange of primers 8 and 8a for primers 8b and 8c (which have an A/A mismatch four bases from their 3'-ends) showed marginal increased specificity (fig. 3 lanes 41 to 46).

DISCUSSION

Interest is increasingly being focused on the mutations in the human genome which produce disease states. The number of such mutations characterised at the DNA sequence level is increasing rapidly and this has been substantially aided by the PCR/direct sequencing approach for the analysis of genomic DNA. We previously reported that PCR followed by direct sequencing was absolutely specific for the diagnosis of AAT deficiency (5). In this communication we present ARMS, a system allowing the direct analysis of any locus of interest and thus generally applicable to any inherited disease provided sufficient sequence data is available. ARMS is simple, rapid and reliable providing the capability for accurate pre- and postnatal diagnosis and a means for heterozygote detection. ARMS is still of benefit even if disease-associated mutations, as yet uncharacterised, are linked to characterised polymorphisms. In such instances the technique will allow detailed haplotype analyses to be performed with a minimal quantity of DNA. Accurate prenatal diagnoses are achievable in a few hours if maternal contamination of the foetal material is avoided. An important practical consideration with this approach (as with other PCR-based strategies) is that it is unnecessary to prepare high quality DNA suitable for restriction enzyme digestion.

A prerequisite of ARMS is the absence of a 3'-exonucleolytic proofreading activity associated with the DNA polymerase employed. The lack of such an exonuclease associated with Taq DNA polymerase has been confirmed here by the successful application of ARMS and independently (23,25). Another requirement in the

application of ARMS is that 3'-OH terminal mismatched primers are refractory to extension by the chosen DNA polymerase. This was not apparent from the work of Tindall and Kunkel (25), since their exonuclease assay required, and did generate, polymerase products from C/A mismatched primer/template complexes. Taq polymerase refractory mismatches have been demonstrated in this work for some mismatched primers. In instances where the mismatch is not refractory to extension (as demonstrated with primer/template G/T, A/C, C/A and T/G mismatches at the AAT Z locus) further deliberate mismatches to destabilise the primer/template complexes render the primers increasingly refractory as the additional mismatch is moved progressively closer to the 3' end.

Empirically, the degree of specificity observed with mismatched primers (and thus the requirement for additional destabilisation), correlates with the mismatch type. C/T, A/A and T/T mismatches (which are all either purine/purine or pyrimidine/pyrimidine mismatches) are considerably more refractory to extension by Taq polymerase than G/T, T/G, A/C or C/A mispairs (all purine/pyrimidine mismatches). We have not yet optimised the position for introduction of the deliberate mismatches, nor the type of mismatch, neither have we examined the effect of deliberate base-pair insertions, deletions or modifications which may also be expected to appropriately destabilise non-refractory primers.

It is likely that any of these approaches to deliberately destabilise the ARMS primers and hence improve specificity may be enhanced by reducing dNTP, magnesium and primer concentrations or simply increasing the amplification annealing/extension temperature. Conversely an increase in concentration of these reagents or a decreased amplification annealing temperature might be expected to have an adverse effect on specificities of primers which previously generated no unwanted product.

We have deliberately chosen relatively unforgiving amplification conditions in this series of experiments so as to challenge the basic concept fully. Removal of tubes from the heating block to facilitate the addition of enzyme after initial DNA denaturation at 100°C was performed in the experiments described herein.

Undoubtedly this allows cooling of reaction mixtures and is difficult to control precisely. Products were generated at the AAT Z locus with mismatched, non-destabilised primers and primers with additional mismatches 7, 5 and 4 bases from their 3'-termini. It is conceivable that the generation of these products will be avoided, if, after heat denaturation of the genomic DNA in the presence of primers the reactions are not allowed to briefly cool during enzyme addition. These products may result from a proportion of template molecules being primed at a less stringent lower temperature than the routine extension temperature. Any extension products so derived would then be correctly paired with the ARMS primer in subsequent cycles, and so generate the observed unwanted products.

Destabilisation of the ARMS primers where necessary such that anomalous products are not generated has been one of our approaches to the development of this technique at this early semi-manual stage. Obviously, further refinement is possible by optimising such variables as magnesium, dNTP or Taq polymerase concentrations and the precise temperature throughout the ARMS cycles. Careful control of the later variable in particular should be achievable with fully automated ARMS instrumentation. Indeed, addition of Taq polymerase to ARMS reaction mixtures at 60°C and ensuring that the reaction temperature never falls below this, may significantly increase reaction specificity and avoid generation of products on mismatched templates.

The AAT Z mutation is caused by a G to A transition immediately preceded by a C residue. The AAT S phenotype results from an A to T mutation. Analysis of single base mutations within coding regions causing human genetic disease shows that 35% of such mutations have occurred within CpG dinucleotides and that over 90% of these were either C to T or G to A transitions (26). Since such mutations would generate the same primer/template mismatches as at the AAT Z locus, it is expected that destabilisation of ARMS primers will be required for at least 30% of ARMS potential applications.

We have chosen to use large (30mer) primers in this assay since this allows the use of high annealing temperatures to improve specificity and reduces the chance of mispriming elsewhere on

genomic DNA. An alternative approach might be to use shorter primers which span a point mutation such that discrimination between 'normal' and 'mutant' loci is achieved by hybridisation of the primers in an allele-specific manner under appropriately stringent conditions. This type of analysis, also of the AAT 2 locus, has been performed by Dermer and Johnson (27). It is important to note however that in this analysis the specificity of the primers was not absolute, the absence of internal controls could conceivably give rise to incorrect diagnoses and hybridisation to blots of the reaction mixtures was required. Other disadvantages would be that different conditions would have to be determined for each locus of interest which would complicate the simultaneous examination of multiple loci. There would also be the danger of the primers priming at loci other than those desired.

It has not escaped our notice that ARMS may have many other applications in medicine and molecular biology. The technique will be useful for the precise typing of infectious pathogens where characteristic strain-specific base changes can be identified. The analysis of oncogene activation is rendered straightforward as is the detection of deletions in DNA. Many further applications can also be envisaged in the research context.

ACKNOWLEDGEMENTS

We thank Dr D. Holland for the preparation of the oligonucleotides used throughout and all the physicians who provided samples for testing.

Patent applications relating to the methods described here are pending.

REFERENCES

1. Southern, E.M. (1975) J. Mol. Biol. 98 503-517.
2. Kan, Y.W. and Dozy, A.M. (1978) Lancet ii 910-912.
3. Solomon, E. and Bodmer, W.F. (1979) Lancet i 923.
4. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239 487-491.
5. Newton, C.R., Kalsheker, N., Graham, A., Powell, S., Gammack, A., Riley, J. and Markham, A.F. (1988) Nucl. Acid. Res. 16 8233-8243.

6. Stoflet, E.S., Koeberl, D.D., Sarkar, G. and Sommer, S.S. (1988) *Science* 239 491-494.
7. Engelke, D.R., Hoener, P.A. and Collins, F.S. (1988) *Proc. Natl. Acad. Sci. USA*. 85 544-548.
8. Bruun Petersen, K., Kølvrå, S., Bolund, L., Bruun Petersen, G., Koch, J. and Gregersen, N. (1988) *Nucl. Acid. Res.* 16 352.
9. Bugawan, T.L., Saiki, R.K., Levenson, C.H., Watson, R.M. and Erlich, H.A. (1988) *Biotechnology* 6 943-947.
10. Chehab, F.F., Doherty, M., Cai, S., Kan, Y.W., Cooper, S. and Rubin, E.M. (1987) *Nature* 329 293-294.
11. Kogan, S.C., Doherty, M. and Gitschier, J. (1987) *New Engl. J. Med* 317 985-990.
12. Levinson, B., Janco, R., Phillips, J. and Gitschier, J. (1987) *Nucl. Acid. Res.* 15 9797-9805.
13. Wong, C., Dowling, C.E., Saiki, R.K., Higuchi, R.G., Erlich, H.A. and Kazazian, H.H. (1987) *Nature* 330 384-386.
14. Connor, B.J., Reyes, A.A., Morin, C. Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) *Proc. Natl. Acad. Sci. USA*. 80 278-282.
15. Orkin, S.H., Markham, A.F. and Kazazian, H.H. (1983) *J. Clin. Invest.* 71 775-779.
16. Feldman, G., Williamson, R., Beaudet, A.L. and O'Brien, W.E. (1988) *Lancet* ii 102.
17. Williams, C., Williamson, R., Coutelle, C., Loeffler, F., Smith, J. and Ivinson, A. (1988) *Lancet* ii 102-103.
18. Wong, C., Antonarakis, S.E., Goff, S.C., Orkin, S.H., Boehm, C.D. and Kazazian, H.H. (1986) *Proc. Natl. Acad. Sci. USA* 83 6529-6532.
19. Estivill, X., Farrall, M., Scambler, P.J., Bell, G.M., Hawley, K.M.F., Lench, N.J., Bates, G.P., Kruyer, H.C., Frederick, P.A., Stanier, P., Watson, E.K., Williamson, R. and Wainwright, B.J. (1987) *Nature* 326 840-845.
20. Farrall, M., Estivill, X. and Williamson, R. (1987) *Lancet* ii 155-157.
21. Nukiwa, T., Brantly, M., Garver, R., Paul, L., Courtney, M., LeCocq, J.P. and Crystal, R.G. (1986) *J. Clin. Invest.* 77 528-537.
22. Ludwig, E.H., Blackhart, B.D., Pierotti, V.R., Caiati, L., Fortier, C., Knott, T., Scott, J., Mahley, R.W., Levy-Wilson, B. and McCarthy, B.J. (1987) *DNA* 6 363-372.
23. Chien, A., Edgar, D.B. and Trela, J.M. (1976) *J. Bacteriol.* 127 1550-1557.
24. Long, G.L., Chandra, T., Woo, S.L.C., Davie, E.W. and Kurachi, K. (1984) *Biochemistry* 23 4828-4837.
25. Tindall, K.R. and Kunkel, T.A. (1988) *Biochemistry* 27 6008-6013.
26. Cooper, D.N. and Youssoufian, H. (1988) *Hum. Genet.* 78 151-155.
27. Dermer, S.J. and Johnson, E.M. (1988) *Lab. Invest.* 59 403-408.

C14

Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms

Theo T. Nikiforov, Robert B. Rendle, Philip Goelet, Yu-Hui Rogers, Michael L. Kotewicz, Stephen Anderson¹, George L. Trainor² and Michael R. Knapp*

Molecular Tool, Inc., Alpha Center, Hopkins Bayview Research Campus, 5210 Eastern Avenue, Baltimore, MD 21224, ¹Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854 and ²Chemical and Physical Sciences, Research and Development, Du Pont Merck Pharmaceutical Company, PO Box 80353, Wilmington, DE 19880, USA

Received June 29, 1994; Revised and Accepted September 7, 1994

ABSTRACT

A new method for typing single nucleotide polymorphisms in DNA is described. In this method, specific fragments of genomic DNA containing the polymorphic site(s) are first amplified by the polymerase chain reaction (PCR) using one regular and one phosphorothioate-modified primer. The double-stranded PCR product is rendered single-stranded by treatment with the enzyme T7 gene 6 exonuclease, and captured onto individual wells of a 96 well polystyrene plate by hybridization to an immobilized oligonucleotide primer. This primer is designed to hybridize to the single-stranded target DNA immediately adjacent from the polymorphic site of interest. Using the Klenow fragment of *E. coli* DNA polymerase I or the modified T7 DNA polymerase (Sequenase), the 3' end of the capture oligonucleotide is extended by one base using a mixture of one biotin-labeled, one fluorescein-labeled, and two unlabeled dideoxynucleoside triphosphates. Antibody conjugates of alkaline phosphatase and horseradish peroxidase are then used to determine the nature of the extended base in an ELISA format. This paper describes biochemical features of this method in detail. A semi-automated version of the method, which we call Genetic Bit Analysis (GBA), is being used on a large scale for the parentage verification of thoroughbred horses using a predetermined set of 26 diallelic polymorphisms in the equine genome.

INTRODUCTION

Mammalian genomes carry numerous single nucleotide polymorphisms (SNPs). On average, two to three polymorphic sites are found per kilobasepair in human genomic DNA (1). Most of these polymorphisms are 'silent' and do not give rise to detectable phenotypes, but an important subset of mutations are associated with heritable diseases such as cystic fibrosis (2),

sickle cell anemia (3), colorectal cancer (4), and retinitis pigmentosa (5, 6).

The wealth of genetic information associated with SNPs can be exploited in a wide variety of applications ranging from the detection of alleles linked to common genetic diseases, to the identification of individuals, to the use of genetic polymorphisms in gene mapping projects. Each of these applications involves the analysis of a large number of samples and will ultimately require rapid, inexpensive, and highly automated methods for typing DNA sequence variants.

Because of the importance of SNPs, a number of methods have been described for their *detection* and *typing*. In general, methods that can be used to discover new mutations can be applied to the typing of those that are already known. These methods include restriction fragment length polymorphism (RFLP) analysis (7), denaturing gradient gel electrophoresis (8), single strand conformation polymorphism (SSCP) detection (9), and chemical or enzymatic mismatch modification assays (10,11). Although powerful, these techniques typically rely on electrophoretic separation to detect the polymorphisms, are relatively labor-intensive, and are difficult to automate.

Approaches for the large-scale typing of known mutations have been described that can be carried out in a nonelectrophoretic mode. Some of these approaches rely on differential hybridization to discriminate the different alleles (12, 13). Other methods, which base the discrimination on an enzymatic reaction, include the oligonucleotide ligation assay (OLA) (14,15), the ligase chain reaction (16), the allele-specific polymerase chain reaction (17,18), and the primer guided nucleotide incorporation assays (19-23). Of these, both the oligonucleotide ligation assay and the primer guided incorporation techniques have been developed to a stage where they can be used for the typing of a relatively large number of samples.

Here, we present a new primer guided genotyping method (Figure 1). The sequence information surrounding the site of variation in the target DNA is used to design an oligonucleotide

*To whom correspondence should be addressed

primer that is complementary to the region immediately adjacent to, but not including, the variable nucleotide site in the target DNA. A single-stranded nucleic acid target molecule is hybridized to this primer immobilized on the polystyrene of a 96-well microplate. The primer is extended by one haptenated dideoxynucleoside triphosphate using a DNA polymerase in the presence of all four chain terminating dideoxynucleoside triphosphates. Novel haptenated ddNTPs allow discrimination of the incorporated nucleotide to be accomplished using standard, enzyme-linked colorimetry.

Our method differs significantly from other primer guided genotyping methods described in the literature. Most importantly, in our method the extension step is carried out in the presence of chain terminating ddNTPs only, and therefore only one nucleotide can be incorporated at the 3' end of the immobilized primer. Second, the immobilization of the primer rather than the template permits the removal of the latter from the reaction mixture following the extension of the primer and thus elimination of all signals that could arise from nonspecific extension at the 3' end of the template. Thirdly, the technique allows the detection of two possible alleles in the same well of a microtiter plate which results in both operational and biochemical advantages.

In this paper, we give a detailed description of this DNA typing method, called Genetic Bit Analysis (GBA). The 'genetic bit' is the term we have adopted for the most elementary form of genetic information, namely a single DNA nucleotide. GBA is a highly flexible method that can be applied, under a standard set of biochemical conditions, to the typing of any nucleic acid polymorphism whose sequence is known. In this paper we focus on the biochemical basis of GBA. Our experience would suggest that features of specificity and convenience inherent in the GBA biochemistry permit the method to become widely used for typing single nucleotide polymorphisms (SNPs) in both research and clinical laboratory applications.

EXPERIMENTAL

Enzymes

Taq DNA polymerase was obtained from Perkin-Elmer. *E. coli* DNA polymerase, Klenow fragment, and T7 gene 6 exonuclease

were purified from recombinant *E. coli* clones containing suitable expression plasmids (unpublished).

Oligonucleotide synthesis

All oligonucleotides were synthesized using standard phosphoramidite chemistry on an Applied Biosystems 392/394 DNA synthesizer, using reagents obtained from Glen Research (Sterling, VA). For the synthesis of phosphorothioate primers, the sulfurizing reagent tetraethylthiuram disulfide (TETD, Applied Biosystems) was used as recommended by the manufacturer. All oligonucleotides were deprotected with concentrated ammonia and desalted using NAP 5 (0.2 μ mol scale synthesis) or NAP 25 (1 μ mol) gel filtration columns (Pharmacia). Oligonucleotides biotinylated at the 5' end were prepared using a biotin phosphoramidite (DMT-Biotin-C6-PA), obtained from Cambridge Research Biochemicals, Inc. (Wilmington, DE). The coupling time of this phosphoramidite was extended to two minutes. The abasic C₃ linker was introduced using the Spacer Phosphoramidite C₃ (Glen Research).

The sequences of the oligonucleotides used in experiments described in this paper are given in Table 1.

Immobilization of oligonucleotides onto 96-well ELISA plates

Immulon 4 plates (Dynatech Laboratories, Chantilly, VA) were used for all experiments shown. Fifty μ l aliquots of a 0.2 μ M oligonucleotide solution in a freshly prepared 20 mM solution of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, obtained from Sigma) in water were added to individual wells of a 96 well plate and incubated overnight at room temperature. The plates were then washed with a solution of TNTw (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween-20). The same procedure was used for other experiments, in which EDC was replaced in the immobilization step by NaCl, tetramethylammonium chloride, cetyltrimethylammonium bromide, and octyldimethylamine hydrochloride.

We have tested a number of different commercially available 96 well plates for their suitability for oligonucleotide immobilization. In general, plates that are described as having a more hydrophilic surface gave good results, whereas those with a hydrophobic surface were found unsuitable. Examples of

Table 1. Sequences of oligonucleotides used in this paper (B denotes a biotin residue; X is a C₃ linker; phosphorothioate bonds are located between the underlined residues)

number	sequence	use
308	5'AGCCTCCGACCGCGTGGTGCCTGGT	GBA primer
308T	5'AGCCTCCTACCGCGTGGTGCCTGGT	GBA primer
308M1	5'AGCCTCCXACCGCGTGGTGCCTGGT	GBA primer
680	5'GAGATGCAGCTCTAAGTCTGTGGG	GBA primer
680T	5'GAGATTCAGCTCTAAGTCTGTGGG	GBA primer
1112	5'AGTATAATAATCAGATATGTTAGC	labeled probe
1676	5'BCCACGGCTAACATACTGTGATTATTACTTAGAT	GBA primer
1464	5'BAATAAGGGGAAACAATTCAGCCCA	synthetic template
501	5'GTTATGGGCTGAATTGTTCCCTAATTT	synthetic template
713	5'TTCTACATTCATTTCTTGTCTGT	synth. template
1302	5'GGAGAACAGAACAGAAAATGAATATGA- ATGTAGAAGCAT	synth. template
1473	5'CCACAACAGAACAGAAAATGAATATGA- ATGTAGAAGCAT	synth. template
1474	5'AACAGAACAGAAAATGAATATGAATGT- AGAAGCAT	synth. template
1214	5' <u>ACCTTCAAA</u> ACTCAACTCAGCTCTT PCR primer	PCR primer
1215	5'TTTACCAATGAGAAGGACATCTAAG	

suitable plates include Immulon 4 (Dynatech); Maxisorp (Nunc); and ImmunoWare plates (Pierce). No attachment could be achieved on Immulon 1 (Dynatech) and Polysorp (Nunc) plates.

DNA isolation and PCR amplification

Horse genomic DNA, isolated from swabs of the nasal mucosa, was the source of DNA in all PCR amplifications. A foam-tipped swab on a six-inch plastic stick was inserted one to two inches into the horse's nostril and immediately immersed in transport buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 5 mM EDTA, 0.5% SDS). The swab remained stored in this solution under ambient conditions until arrival at the laboratory. DNA was isolated from this mixture by treatment with a mixture of guanidine hydrochloride and ethanol and adsorption to glass matrices (e.g., MagicTM resin, obtained from Promega, Madison, WI), followed by recovery in 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA. PCRs were carried out in a total volume of 30–50 μ l. The final concentration of the PCR primers was 0.5 μ M. Following an initial two minute denaturation step at 95°C, thirty-five cycles were carried out, each consisting of denaturation (1 min at 95°C), annealing (2 min at 60°C), and extension (3 minutes at 72°C). *Taq* DNA polymerase was used at a concentration of 0.025 units/ μ l. The final composition of the PCR buffer was: 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, and 170 μ g/ml BSA.

Preparation of single-stranded PCR fragments

Single-stranded DNA was prepared from double-stranded PCR products as described (25). One of the strands was protected from exonuclease hydrolysis by the introduction, during synthesis, of four phosphorothioate groups at the 5' end of one of each pair of the PCR primers. Following the PCR amplification, T7 gene 6 exonuclease, diluted in a buffer containing 50 mM Tris-HCl, pH 7.5, 1 mM DTT, and 100 μ g/ml BSA, was added to a final concentration of 0.5 units/ μ l. Incubation with this enzyme was for one hour at room temperature.

Hybridization of single-stranded PCR fragments to oligonucleotides immobilized onto ELISA plates

After the exonuclease treatment, an equal volume of 3 M NaCl, 20 mM EDTA was added to the reaction mixture and 20 μ l aliquots of the resulting solution transferred to individual wells containing the appropriate immobilized oligonucleotide molecule. Hybridization was carried out for 30 min at room temperature and was followed by washing with TNTw.

Labeled dideoxynucleoside triphosphates

All biotin- and fluorescein-labeled chain-terminating 2',3'-dideoxynucleoside triphosphates used in the single nucleotide extension reaction were purchased from Du Pont NEN, (Wilmington, DE). A selection of labeled ddNTPs are commercially available from that supplier (sold as Renaissance non-radioactive products). These compounds are derivatives of amino-propynyl-substituted 2',3'-dideoxypyrimidines or 2',3'-dideoxy-7-deazapurines. The chemistry of these chain terminators and their use in DNA sequencing have been described (24).

Solid-phase primer extension

Following the hybridization step, 20 μ l of polymerase extension mix was added to each well. The extension mix contained 20 mM Tris-HCl, pH 7.5; 10 mM MgCl₂; 25 mM NaCl; 10 mM MnCl₂; 15 mM sodium isocitrate; 1.5 μ M of two unlabeled

2',3'-dideoxynucleoside triphosphates; 1.5 μ M of one biotin-labeled and 1.5 μ M of one fluorescein-labeled 2',3'-dideoxynucleoside triphosphate; and the Klenow fragment of *E. coli* DNA polymerase I (0.3 units per well). The polymerase was diluted in a buffer containing 10 mM Tris-HCl, pH 7.5, 5 mM DTT, and 0.5 mg/ml BSA. The extension reaction was carried out for 10 min at room temperature. The plates were subsequently washed once with TNTw, once with 0.2 N NaOH, and three additional times with TNTw.

Colorimetric detection of the incorporated nucleotides

After the extension step, the wells were incubated for 30 min at room temperature with 40 μ l of 1% BSA in TNTw containing an alkaline phosphatase conjugate of anti-fluorescein (Bioscience International, Kennebunk, ME) and a horseradish peroxidase-conjugated anti-biotin (Vector Laboratories, Burlingame, CA). The dilution factor was 1:500 for the anti-fluorescein and 1:1000

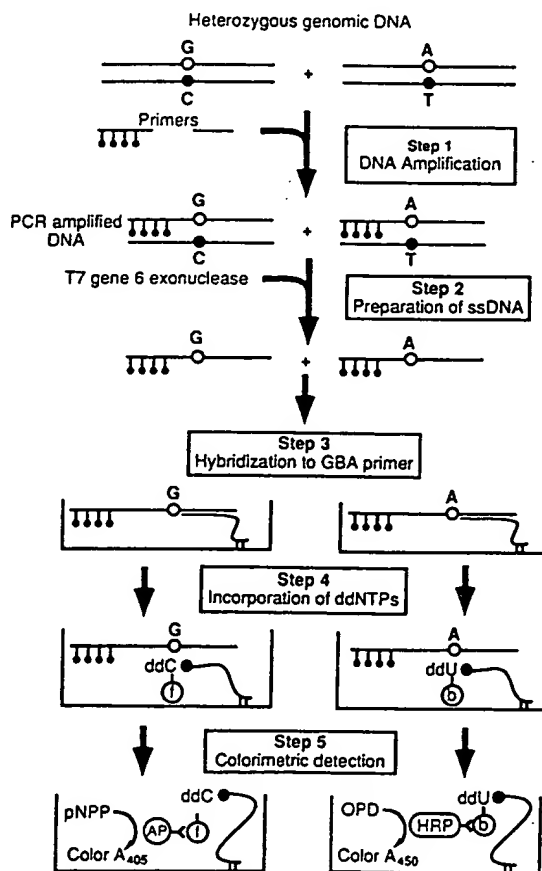


Figure 1. Schematic representation of the individual steps of single nucleotide typing by GBA. In Step 1, a DNA fragment containing the polymorphic site to be typed is amplified by PCR using a primer containing four phosphorothioate bonds at the 5' end. In Step 2, the double-stranded PCR product is rendered single-stranded by treatment with T7 gene 6 exonuclease. In Step 3, the single-stranded DNA template is captured by hybridization to a primer immobilized to the wells of a microtiter plate, whereby the polymorphic site of the template is located immediately downstream from the 3' end of the primer. In Step 4, the 3' end of the primer is enzymatically extended by one nucleotide using haptenated ddNTPs. In Step 5, the nature of the incorporated nucleotide(s) is revealed by an enzyme-linked assay.

for the antibiotin. These dilutions were calibrated to give colorimetric signals of approximately equal intensity with the two alleles.

After washing, the presence of alkaline phosphatase was detected first by addition of 100 μ l per well of a 1.5 mg/ml solution of p-nitrophenyl phosphate in 100 mM diethanolamine, pH 9.5, 20 mM $MgCl_2$. The plate was immediately placed in a kinetic microplate reader (Molecular Devices Corporation, Menlo Park, CA), and the development of color was followed at 405 nm for 2 min. The results were expressed as mOD_{405}/min . The plates were then washed again and incubated with 100 μ l of a 1 mg/ml solution of o-phenylenediamine in 0.1 M citric acid, pH 4.5, containing 0.012% H_2O_2 . The reaction was followed by measuring the change of light absorbance at 450 nm as above. Most experiments in microtiter plates described in this article have been carried out at least in triplicate, and the results presented are the averaged numbers. The two enzymes used, alkaline phosphatase and horseradish peroxidase have significantly different pH optima (9.8 vs. 4.5), thus, if endpoint readings are to be taken, it is preferable that the incubations with the two antibody conjugates are carried out sequentially rather than simultaneously in order to avoid partial inactivation of the second antibody conjugate.

Single nucleotide primer extension in solution

A solution containing 2 pmole of the synthetic template # 501 (see Table 1 for oligonucleotide sequences) and 800 fmole of the 5' biotinylated primer # 1464 in 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 25 mM NaCl, 10 mM $MnCl_2$ and 15 mM sodium isocitrate, was heated to 95°C for 10 min, then slowly cooled down to room temperature to anneal the primer to the template. Aliquots of this solution were then added to four individual tubes, each containing solutions of one fluorescein-labeled ddNTP, the three other unlabeled ddNTPs, and the Klenow polymerase. After 10 min incubation at room temperature, aliquots of these mixtures were transferred to individual wells of an avidin-coated microtiter plate to capture the extension complexes via the 5' biotin residue of the primer. The wells were then washed with 0.2 N NaOH to remove the template strand and the presence of fluorescein was detected using an anti-fluorescein HRP conjugate.

RESULTS AND DISCUSSION

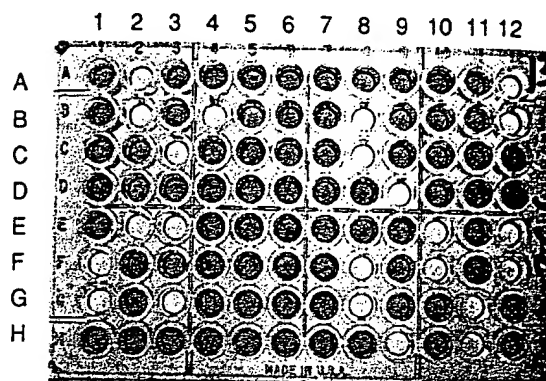
DNA typing by Genetic Bit Analysis (GBA)

The individual steps of GBA are shown schematically in Figure 1. We have developed a test for the parentage verification of thoroughbred horses based on GBA, whereby each horse is typed at 26 different, diallelic loci. The use of 96 well microtiter plates has allowed us to develop a semi-automated version of the test, taking advantage of a number of commercially available, automated liquid handlers for that format. In this automated version of the test, 88 horses are typed, together with suitable controls, with respect to one locus on one microtiter plate. Figure 2 shows a typical result from such a test. This Figure represents the results from the typing of 88 horses with respect to locus JH261-1, a single nucleotide polymorphism present in the equine genome (manuscript in preparation). The same microtiter plate was photographed after development of the colorimetric reaction for alkaline phosphatase which reveals allele 1 (incorporation of fluoresceinated ddCTP, top) and, with appropriate processing,

after the colorimetric reaction for horseradish peroxidase that reveals allele 2 (incorporation of biotinylated ddUTP, bottom). Controls for specific and non-specific effects were also run (see legend to Figure 2). Genotypes are visually scorable: CC homozygotes give a strong reaction with alkaline phosphatase but are negative for horseradish peroxidase, TT homozygotes have the opposite profile, and CT heterozygotes are positive for both enzymatic reactions.

Absorbance values for a comparable set of 88 horses typed with respect to the same polymorphic locus JH261-1 were

A.



B.

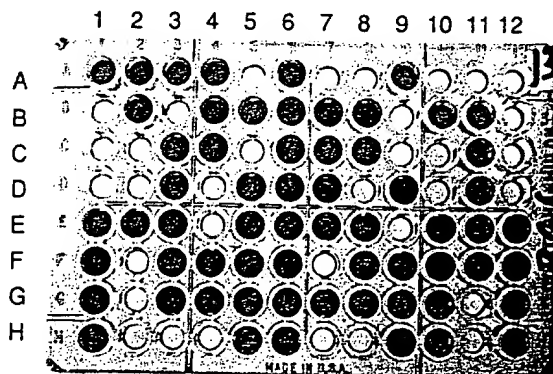


Figure 2. Colorimetric detection of two alleles on a single microtiter plate. The GBA primer # 1112 (see text for sequence) was immobilized in all wells of this plate using EDC. A 116 bp fragment was amplified from the genomic DNA of 86 different thoroughbred horses, using the primers # 1214 and # 1215. This PCR fragment contains a CT diallelic polymorphism (JH261). Following the hybridization of the single-stranded PCR templates to the GBA primer, an extension reaction was carried out using fluorescein-labeled ddCTP and biotin-labeled ddUTP. The two haptens were detected as described in the text. A. detection of ddCTP incorporation using an alkaline phosphatase conjugate; B. detection of ddUTP incorporation using a horseradish peroxidase conjugate. The plate contained the following controls: a) no template was added to wells A12 and B12 (template-independent extension controls); b) a 35 mer synthetic template (250 fmol) giving incorporation of a ddCTP was added to wells C12 and D12; c) a similar synthetic template giving incorporation of a ddUTP was added to wells E12 and F12; d) a mixture of both templates was added to wells G12 and H12; e) to control for PCR crosscontamination, negative PCR reactions were carried out and added to wells G11 and H11.

measured in a 96-well spectrophotometer. The graph in Figure 3 depicts the results quantitatively as a scatter plot. The values for each horse by typing for allele 1 (C) are indicated on the X-axis; those for allele 2 (T) on the Y-axis. Previous experiments which typed this locus with respect to several hundred horses failed to find a third allele (data not shown). The data in Figure 3 are consistent with JH261-1 being a diallelic, single nucleotide polymorphism (SNP).

The genotype groups are circled in Figure 3. A summary of the quantitative data is given in Table 2. It can be seen that the mean values for the two alleles can be calibrated to be roughly equivalent. Furthermore, variability between horses is surprisingly small. In the experiment shown, eight test samples produced signals with both alleles that were judged unscorable. Theoretically, this result could have been produced because of a failure in one or more of the biochemical reactions leading to the colorimetric data, because of a failure to amplify due to allelic variability in the primer sites, because of an allelic variability in the GBA primer site, or because the horses in question possessed an allele other than A or G in the template strand. We have investigated a large number of these results further and in all cases thus far examined, biochemical failure is the explanation of failure. This has been shown by analysis of the PCR reactions by gel electrophoresis to be PCR failure in most cases. The failure rate has been found to be substantially lower when a standardized control horse genomic DNA is used. For this reason, we believe that variability in sample quality produces most failure in our system. However, it can be anticipated that situations leading to failures of the other three types will arise in complex genomes and therefore adequate characterization of variability in target genetic loci is required for optimal utilization of the GBA method.

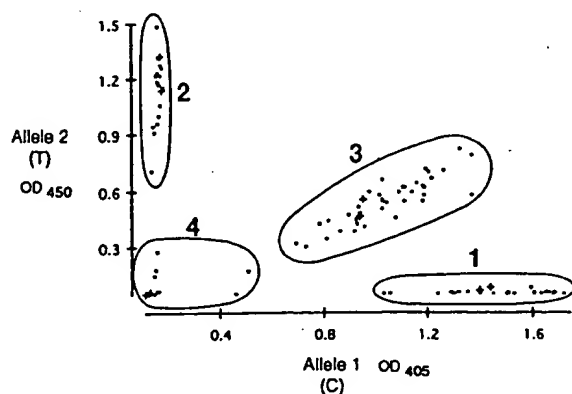


Figure 3. Scatter plot representation of the results from typing of 88 different thoroughbred horses with respect to the diallelic polymorphic locus JH261-1. The results shown are those from endpoint readings of the plate taken after 24 min of incubation with the enzyme substrates for the two antibody conjugates. Horse samples are indicated by dots while the control wells are plotted with '+'s. The points fall into four categories: 1) high values for C and low values for T; 2) high values for T and low values for C; 3) high values for both alleles; 4) high values for neither allele. The controls run in duplicate were: a) synthetic oligonucleotide molecules which mimic the PCR template for this locus and which possess an A at the variable position (+ 's found in group #1); b) which possess an A at the variable position (+ 's found in group #2); c) a mixture of these synthetic oligonucleotide templates (+ 's found in group #3); and d) PCR reactions to which no horse genomic DNA had been added as the amplification template (+ 's found in group #4).

Immobilization of the GBA primer to microtiter plates

We have previously described the details of our method for oligonucleotide immobilization onto polystyrene plates (25). Briefly, the method consists of incubating the oligonucleotide on the microtiter plates in a dilute solution of an organic or inorganic salt, followed by washing with a solution containing 0.05% Tween 20. We have shown that the oligonucleotides immobilized in this way are capable of specific hybridization to complementary templates and have used these findings to develop a convenient microplate-based PCR product detection assay. Other authors have also successfully immobilized oligonucleotide probes to the surface of polystyrene plates using NaCl-containing buffers and used those in hybridization-based assays (26).

In the current experiments, we have tested different compounds in the immobilization process and found that a number of chemically divergent reagents are capable of promoting this process. For example, the efficiency of immobilization of oligonucleotide #1112 using cetyltrimethylammonium bromide (CTAB) and tetramethylammonium chloride (TMAC) was compared (Figure 4). As negative controls, the oligonucleotides were added to some of the wells as aqueous solutions, without any immobilization reagents. To assess the immobilization process, following an overnight incubation with the immobilization reagents, the biotinylated oligonucleotide #1676, which is complementary to #1112, was added to the wells of the microtiter plate at a range of different concentrations. The amount of this biotinylated probe captured by hybridization to the immobilized oligonucleotide #1112 was determined by an enzyme-linked assay for the biotin residue. The results of this experiment are represented graphically in Fig. 4.

The results shown in Fig. 4 as well as similar results obtained with other compounds suggest that the immobilization reagents can be divided in two groups. The first group consists of chemicals like NaCl and TMAC, which work best when used at relatively high concentrations, generally higher than 50 mM, and best at 250 to 500 mM. Even concentrations as high as 1 M can be used without any noticeable adverse effect on the immobilization. The second group of immobilization reagents consists of chemicals that are characterized by the presence of two structural features: a positively charged 'head' and a relatively hydrophobic 'tail'. These are the typical features of cationic detergents. Representatives of this group are the cationic detergent cetyltrimethyl ammonium bromide (CTAB), octyldimethylamine hydrochloride, and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC). These compounds can be used for oligonucleotide immobilization at very low concentrations, as low as 0.03 mM for CTAB, but a lower hybridization signal is obtained when they are used at higher concentrations. The

Table 2. Typing of a single nucleotide polymorphism in equine DNA (locus JH261-1)

Genotype	Value AP	(SD)	Value HRP	(SD)
CC	1.478	0.190	0.051	0.009
CT	1.044	0.165	0.546	0.121
TT	0.155	0.012	1.143	0.186
NS	0.194	0.140	0.097	0.075

Endpoint readings were taken after 24 min of incubation with the colorimetric substrates. The average signals (in OD units at 405 and 450 nm) and corresponding standard deviations (SD) for 88 different horses are shown. (NS, no signal).

inhibitory concentrations differ among the reagents of this group. For CTAB, it is as low as 0.5 mM, whereas for EDC it is about 500 mM. It should be noted that the critical micelle concentration, CMC, for CTAB is about 1 mM. Thus, it is possible that once micelles are formed, the immobilization is inhibited. Compounds of a similar structure, but with a negatively charged 'head' (e.g., SDS) are completely inactive as oligonucleotide immobilization reagents, as are nonionic detergents.

The mechanism of immobilization in the presence of EDC or cationic detergents is probably very similar to the mechanism of transfer of nucleic acids and proteins through an organic phase in the presence of detergents described recently (27).

In another experiment, radioactively labeled oligonucleotides were immobilized to polystyrene plates using EDC. The amount of immobilized oligonucleotide was then determined by counting the amount of radioactivity released upon dissolving the wells in toluene. It was thus determined that approximately 1 pmole of oligonucleotide is immobilized in each well, which corresponds to about 10% of the input oligonucleotide (10 pmole). We also found that the input of oligonucleotide can be reduced to about 3 pmole per well before there is a noticeable decrease in the amount of immobilized material.

PCR amplification, generation of single-stranded DNA templates and their capture by hybridization to the GBA primers

PCR normally produces double-stranded products which do not hybridize to the immobilized capture oligonucleotide without prior strand separation. This strand separation can be achieved by treatment with heat or alkali, but we found the efficiency of hybridization to be low even with such a denaturation step. Asymmetric PCR has also been used for the generation of single-stranded products (28). Unfortunately, asymmetric PCR generates single-stranded products only linearly, and we found the results to be variable. Previously, we have reported a new and efficient method for the generation of single-stranded PCR products following a regular exponential amplification (25). The method is based on the selective protection of one of the strands of the PCR product from enzymatic hydrolysis of T7 gene 6 exonuclease

by the incorporation of four phosphorothioate bonds into the 5' end of that strand using modified PCR primers. The exonuclease method generates single-stranded products with high efficiency, and they are ideally suited for the subsequent hybridization to the immobilized primer.

The optimal length of the GBA primers immobilized in the microtiter plates appears to be between 20 and 25 bases. Oligonucleotides shorter than 20 bases usually give lower signals, and virtually no signals are seen with primers shorter than 10 bases. Apparently, parts of the immobilized oligonucleotides are inaccessible for hybridization because they are involved with interactions with the solid phase. This is supported by the finding that some 'hybrid' 25 mer primers that contain only about 12 to 15 bases at the 3' end exactly matching the template give signals as strong as those seen with completely matching 25 mers. Primers longer than 30 bases produce only slightly better extension signals, but tend to be more prone to template-independent extension (see below).

The solid phase primer extension reaction

In the enzymatic primer extension step, a single dideoxynucleoside triphosphate is incorporated at the 3' end of the immobilized GBA primer. The nature of the nucleotide at the polymorphic site of the template determines which of the four ddNTPs contained in the extension mixture will be incorporated by the polymerase. We have found that both the modified T7 DNA polymerase (Sequenase) and the Klenow fragment of *E. coli* DNA polymerase I are suitable enzymes for the primer extension. Both polymerases assure a very high signal-to-noise ratio. The Klenow polymerase possesses a 3'–5' exonucleolytic activity which Sequenase lacks (29). This exonuclease activity did not cause problems in the template-directed extension, but we have encountered cases of template-independent extension that are the result of this activity (see below).

During the development of GBA, we have found that false positive results can be generated by three different mechanisms. The first of these is trivial, and consists of self-extension at the 3' end of the template DNA at the same time as the 3' end of the immobilized GBA primer is being extended. This source of

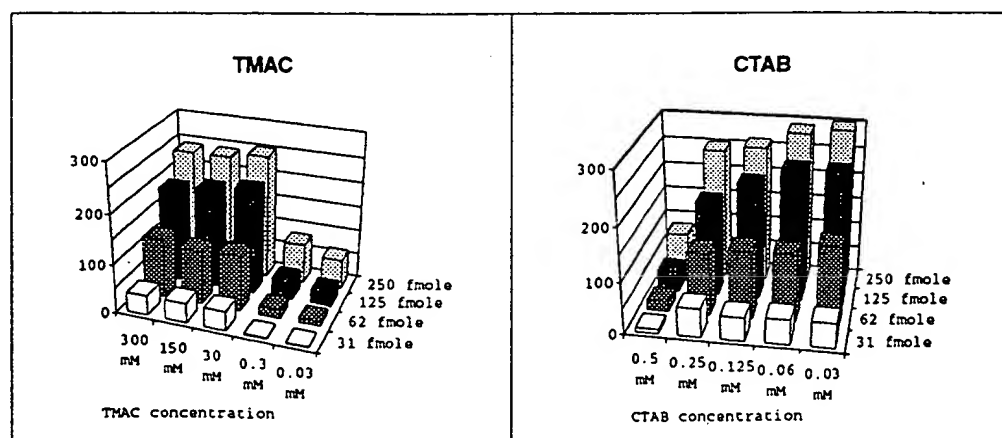


Figure 4. Comparison of CTAB and TMAC as oligonucleotide immobilization reagents. Oligonucleotide #1112 was immobilized to Immulon 4 plates using varying concentrations of these two reagents, and then hybridized to the complementary biotinylated oligonucleotide #1676, used at concentrations of 31, 62, 125, and 250 fmole per well. The signals obtained in the colorimetric assay are given in mOD₄₅₀/min.

'noise' is eliminated simply by briefly washing the plates after the polymerase extension step with a 0.2 N NaOH solution.

The second source of false positive signal, template-independent extension of the GBA primers, is the most likely problem to be encountered during the development of a GBA for a new polymorphism. This is the result of the formation of inter- or intramolecular secondary structures by the immobilized GBA primers and their subsequent enzymatic extension. Table 3 summarizes the results seen in the template-independent extension of two GBA primers, #308 and #680, as well as some of their modified versions.

A typical example of a primer showing template-independent extension is oligonucleotide #308. In this experiment, four separate wells were used to characterize the extension reaction. In each, only one of the ddNTPs was labeled with biotin while the other three were unlabeled. When this primer was immobilized on a plate, the extension reaction produced a strong signal due to incorporation of C. Analysis of the sequence of this oligonucleotide shows that it might be able to form relatively stable inter- or intramolecular partially self-complementary structures. These are shown in Figure 5. In both structures, the highlighted G residue will dictate the incorporation of a C by the polymerase.

To test whether these structures could explain the template-independent signal, a modified version of this primer was synthesized where the G residue of the original sequence was replaced by a T. This modified primer, #308T showed a strong template-independent extension signal in A. In another modified version of the same oligonucleotide, #308M1, the G residue of the original sequence was replaced by the abasic C₃ linker

Table 3. Template-independent extension of primers # 308, 680, and their modified versions

primer #	base G	base A	base T	base C
308	0.2	1.3	0.1	21.5
308T	2.0	48.5	0.2	0.2
308M1	0.5	0.3	0.9	0.5
680	0.4	1.2	0.5	65.5
680T	0.2	58.0	0.7	1.0

Extension reactions were done with the Klenow polymerase. The signals for the four different bases are given in mOD₄₅₀/min.

#308:

5' AGCCTCCGACCGCGTG
 III I G
 TGGTCCGT

5' AGCCTCCGACCGCGTGGTGCCTGGT
 III I I III
 TGGTCCGTGGTGCCTGAGCCTCCGA 5'

#680:

5' GAGATGCAGCTCTAAGTCTGTGGG
 IIII II II IIII
 GCGTGTCTGTAATCTCGACGTAGAG 5'

Figure 5. Postulated secondary structures of two GBA primers that could lead to the observed template-independent extensions. The nature of the incorporated nucleotides will be determined by the highlighted bases.

OPO₃CH₂CH₂CH₂ using a commercially available phosphoramidite. The modified primer #308M1 did not show any template-independent extension. The presence of the abasic linker within the sequence of 308M1 did not affect its hybridization and extension to synthetic or PCR-derived templates.

Primer #680 is an example of template-independent extension that is dependent on the polymerase used for extension. With this primer, a strong template-independent extension by a C was seen only when the extension was carried out with the Klenow polymerase (see Table 3), but not when the extension was carried out with the modified T7 DNA polymerase (Sequenase). It is reasonable to assume that this oligonucleotide forms the self-complementary structure shown in Figure 5. The four non-base paired residues are cleaved off by the 3'–5' exonuclease of the Klenow polymerase which then inserts a C opposite the highlighted G. When we replaced this deoxyguanosine residue by a thymidine in the modified primer #680T, the template-dependent noise was changed to an A, as expected (Table 3). Because of the possibility of template-independent extension of the GBA primers, each new GBA primer should be tested for this type of extension before being used in typing experiments. This phenomenon would produce inappropriate typings in some percentage of cases especially when PCR yield has been low. Commercially available computer programs for DNA analysis (e.g., 'Oligo', National Biosciences, Inc., Plymouth, MN) can be used to predict potential secondary structures. The required modifications can be incorporated in the sequence of the appropriate GBA primer.

The third source of false positive signals is template-dependent, but in contrast to the first type of 'noise' described above, it is the result of extension at the 3' end of the GBA primer and not the template. Signals of variable strength are sometimes generated in one of the 'wrong' bases, i.e., a base that is not consistent with the sequence of the template to be typed. We have observed the same 'noise' profile for both polymerases tested, Sequenase and Klenow. This type of 'noise' is notably dependent on the amount of template molecules hybridized to the GBA primer and can be especially serious when high concentrations (usually, more than 500 fmole) of synthetic template molecules are used in GBA. In the majority of cases when PCR generated templates are typed, this type of noise is undetectable or very weak (signal-to-noise ratios: >20), but on rare occasions can be strong enough to cause false interpretation of the genotyping results. A summary of the results of some GBA experiments with template-dependent 'noise' is given in Table 4.

The biochemical basis of this type of 'noise' is uncertain, but it may be the result of mishybridization of the GBA primer to

Table 4. Analysis of template-dependent noise

GBA experiment	base G	base A	base T	base C
713+1302	12.0	35.1	105.4	35.5
713+1473	60.4	50.8	160.0	2.5
713+1474	23.4	64.7	120.4	2.5
501+1464	5.5	12.4	85.0	1.0
(in solution)				
501+1464	7.0	35.5	170.2	2.0
(solid phase)				

Signals are given in mOD₄₅₀/min. 500 fmole of the synthetic templates was used in the solid phase extension experiments involving primer #713. Incorporation of a labeled T was expected in all of these experiments.

the template and/or misincorporation of the wrong ddNTP by the polymerase. Indeed, the polymerases used may in some sequence contexts display a higher misincorporation rate with the labeled ddNTPs used. We have analyzed in more detail the template-dependent 'noise' observed with the GBA primer #713 and the synthetic template #1302 (see Table 4). In GBA, this template-primer combination gives significant template-dependent 'noise' in bases C and in A. We synthesized and tested two modified templates, #1473 and #1474, which differ from #1302 only by a few bases at the 5' end, i.e., in a part of the template that is not expected to hybridize to the GBA primer. In oligonucleotide #1473, three deoxyguanosine residues of #1302 are changed to deoxycytidines; in oligonucleotide #1474, the part of the synthetic template extending beyond the GBA primer is reduced to one single deoxyadenosine. The results shown in Table 4 demonstrate that the template-dependent 'noise' is influenced by the sequence surrounding the residue of the template DNA that directs the dideoxynucleotide incorporation. Thus, the replacement of the deoxyguanosines of #1302 with deoxycytidines in #1473, or their elimination in #1474, reduces the 'noise' in C and, in the case of template #1473, increases the noise in G. Such context-dependent effects on the fidelity of DNA polymerases have been reported before (30).

To verify that this 'noise' is not due to the fact that the hybridization and extension reactions are carried out on the solid phase, an extension experiment was carried out in solution (see Experimental). The 5' biotinylated oligonucleotide #1464 was used as a primer and annealed to the oligonucleotide #501. The expected signal in this template-primer combination is a T. In parallel, this typing experiment was carried out on the polystyrene solid phase, by immobilizing the primer #501 and using 500 fmole of oligonucleotide #1464 as the template. These two experiments produced a remarkably similar signal-to-noise profiles (see Table 4). Analogous results were obtained with other primer-template combinations (not shown).

Although we have not found a general solution for eliminating the rare occurrence of template-dependent 'noise', the following approaches have been found to reduce or eliminate the problem. Hybrid GBA primers of the type 5' $X_{12}N_{13}$, where each X position contains equal amounts of the four bases whereas the 13 bases at the 3' end match exactly the template were sometimes found to give better signal-to-noise ratios than completely matching 25 mers. This could be due to reduced or eliminated mishybridization with these primers. The signal-to-noise ratio was also improved by performing the extension reaction at 5°C rather than at room temperature, and also by decreasing the concentration of all ddNTPs in this step. These two factors probably affect the fidelity of the polymerase. Finally, this 'noise' can usually be avoided by switching the primer protected from exonuclease digestion and typing the same polymorphism on the opposite DNA strand with a suitable GBA primer.

Colorimetric detection of the incorporated labeled nucleotide

As shown above, it is possible to type single nucleotide polymorphisms by GBA by including only one labeled ddNTP per well. However, the use of two labeled ddNTPs allows the determination of both alleles in a diallelic locus to be carried out in the same well. This not only reduces the amount of PCR generated template required and results in considerable savings of labeled chain terminators, but serves as a very useful internal control for all post-PCR steps. For example, for a particular

template, a blank well in the one-base-per-well mode could be due to homozygosity of the other type but also to failure of one of the post-PCR steps (hybridization, extension). In the two-bases-per-well mode, lack of signal for one of the bases can only be due to homozygosity of the other type or failure of the enzyme-linked assay for this allele. The latter hypothesis can be routinely excluded with suitable controls.

To date, we have found no randomly discovered site of single nucleotide polymorphism to be tri- or tetra-allelic. This is consistent with findings in other laboratories (Deborah Nickerson, personal communication). However, if such a site were to be encountered, it could be typed using a second well whereby incorporation of the other two nucleotides could be examined through inclusion of their ddNTP derivatives. Alternatively, other haptenated ddNTPs could be used simultaneously assuming an appropriate antibody-enzyme conjugate was available. Failure to test for the alleles which are rare in the population can produce incompatible data in legitimate pedigrees because heterozygotes where one chromosome possesses a 'null' allele would appear to be a homozygote for one of the tested alleles. A falsely typed heterozygous parent would appear to be excluded if its offspring inherits the 'null' allele. This explanation can be considered likely if many single-locus exclusions are observed for a particular locus in a panel of markers used for genetic studies.

CONCLUSION

Compelling arguments exist for the development of DNA-based assays, which can be performed on a very large scale, for the analysis of known polymorphisms in complex genomes. In this paper we give biochemical details concerning a new genotyping procedure, GBA (genetic bit analysis), that is simple, convenient, and automatable. In this method, sequence-specific primer annealing is used to select a unique polymorphic site in a nucleic acid sample, and interrogation of this site is accomplished via the highly accurate DNA polymerase reaction using a set of novel, commercially available, non-radioactive dideoxynucleotide analogs.

An important feature of GBA is that it does not rely on nucleic acid hybridization for purposes of nucleotide discrimination. This is in marked contrast to methods such as allele-specific hybridization (12). Rather, nucleic acid hybridization is used only to position the template to be typed with respect to a primer molecule. Nucleotide discrimination is accomplished using a DNA polymerase, an enzyme that has evolved to perform that role. As a result, all GBA reactions can be performed under a standard set of conditions that have been optimized for high throughput operations. Many loci can be tested simultaneously, and new tests can be developed rapidly.

GBA was developed to be a method that can be applied on a very large scale using commercial liquid handling devices. In our parentage verification assay for thoroughbred horses, as many as 180 horses are typed by GBA at 26 diallelic loci by a single technician in a single day. However, GBA can easily be carried out manually. The signals generated are usually strong enough to allow visual interpretation of the results (Figure 2), without the need for a spectrophotometer. Alternatively, data can be acquired on a large scale using high-volume, stacking plate readers. Therefore, it has the potential to become a method for the typing of single nucleotide polymorphisms across a broad spectrum of applications.

ACKNOWLEDGEMENTS

Special thanks are due to Dr Jonathan Karn for his advice and continuous support of this project. We are grateful to Dr Sidney Brenner for stimulating discussions. The technical assistance of Dr Florence McAuliffe and Mr Richard Douglas in the early stages of this work is also gratefully acknowledged.

REFERENCES

- Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S. and Schmidtke, J. (1985) *Hum. Genet.* 69, 201-205.
- Kerem, B.-S., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L.-C. (1989) *Science* 245, 1073-1080.
- Watson, J.T., Marotta, C.A., Forget, B.G. and Weissman, S.M. (1977) *Trans. Assoc. Am. Physicians* 90, 117-126.
- Fearon, E.R. and Vogelstein, B. (1990) *Cell* 61, 759-767.
- Sung, C.H., Davenport, C.M., Hennessey, J.C., Maumence, I.H., Jacobson, S.G., Heckenlively, J.R., Nowakowski, R., Fishman, G., Gouras, P. and Nathans, J. (1991) *Proc. Natl. Acad. Sci. USA* 88, 6481-6485.
- Dryja, T.P., Hahn, L.B., Cowley, G.S., McGee, T.L. and Berson, E.L. (1991) *Proc. Natl. Acad. Sci. USA* 88, 9370-9374.
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) *Am. J. Hum. Genet.* 32, 314-331.
- Myers, R.M., Maniatis, T. and Lerman, L.S., (1987) *Methods Enzymol.* 155, 501-527.
- Hayashi, K. (1991) *PCR Methods Applic.* 1, 34-38.
- Meyers, R.M., Larin, Z. and Maniatis, T. (1985) *Science* 230, 1242-1246.
- Cotton, R.G.H., Rodrigues, N.R. and Campbell, R.D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4397-4401.
- Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) *Proc. Natl. Acad. Sci. USA* 80, 278-282.
- Saiki, R.K., Walsh, P.S., Levenson, C.H. and Erlich, H.A. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6230-6234.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) *Science*, 241, 1077-1080.
- Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L. and Landegren, U. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8923-8927.
- Barany, F. (1991) *Proc. Natl. Acad. Sci. USA* 88, 189-193.
- Wu, D.Y., Ugozzoli, L., Pal, B.K. and Wallace, R.B. (1989) *Proc. Natl. Acad. Sci. USA* 86, 2757-2760.
- Newton, C.R., Graham, A., Heptinstall, L. E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J. C. and Markham, A.F. (1989) *Nucleic Acids Res.* 17, 2503-2516.
- Syvänen, A.-C., Aalto-Setälä, K., Harju, L., Kontula, K. and Soderlund, H. (1990) *Genomics* 8, 684-692.
- Kuppuswami, M.N., Hoffmann, J.W., Kasper, C.K., Spitzer, S.G., Groce, S.L. and Bajaj, S.P. (1991) *Proc. Natl. Acad. Sci. USA* 88, 1143-1147.
- Nyren, P., Pettersson, B. and Uhlen, M. (1993) *Anal. Biochem.* 208, 171-175.
- Ugozzoli, L., Wahlquist, J.M., Ehsani, A., Kaplan, B.E. and Wallace, R.B. (1992) *GA7A* 9, 107-112.
- Syvänen, A.-C., Sajantila, A. and Lukka, M. (1993) *Am. J. Hum. Genet.* 52, 46-59.
- Prober, J.M., Trainor, G.L., Dam, R.D., Hobbs, F.W., Robertson, C.W., Zagurski, R.J., Cocuzza, A.J., Jensen, M. and Baumeister, K. (1987) *Science* 238, 336-341.
- Nikiforov, T.T., Rendle, R.B., Kotewicz, M.L. and Rogers, Y.-H. (1994) *PCR Methods Applic.* 3, 285-291.
- Balaguer, P., Térouanne, B., Allibert, P., Cros, P., Boussioux, A.-M., Mandrand, B. and Nicolas, J.-C. (1993) *Mol. Cell. Probes* 7, 155-159.
- Bromberg, L.E. and Klibanov, A.M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 143-147.
- Gyllenstein, U. and Erlich, H.A. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7652-7656.
- Tabor, S. and Richardson, CC. (1989) *J. Biol. Chem.* 264, 6447-6458.
- Goodman, M.F., Creighton, S., Bloom, L.B. and Petruska, J. (1993) *Crit. Rev. Biochem. Mol. Biol.* 28, 83-126.

Rapid and Sensitive Detection of Point Mutations and DNA Polymorphisms Using the Polymerase Chain Reaction

MASATO ORITA, YUICHI SUZUKI, TAKAO SEKIYA, AND KENSHI HAYASHI¹

Oncogene Division, National Cancer Center Research Institute, 1-1, Tsukiji 5-chome, Chuo-ku, Tokyo, Japan

Received April 24, 1989, revised August 1, 1989

We report a rapid and sensitive method for the detection of base changes in given sequences of genomic DNA. This technique is based on the facts that specific regions of genomic sequences can be efficiently labeled and amplified simultaneously by using labeled substrates in the polymerase chain reaction and that in nondenaturing polyacrylamide gels, the electrophoretic mobility of single-stranded nucleic acid depends not only on its size but also on its sequence. The process does not involve restriction enzyme digestion, blotting, or hybridization to probes. We found that most single base changes in up to 200-base fragments could be detected as mobility shifts. *RAS* oncogene activation was detected by this technique. We also show that the interspersed repetitive sequences of human *Alu* repeats are highly polymorphic. © 1989 Academic Press, Inc.

INTRODUCTION

Today, a large volume of DNA sequence data is available, and detection of possible sequence deviations in many individuals or cells is sometimes desired in such cases as search and linkage analyses of genetic polymorphisms, detection of DNA aberrations in cancer tissues, and searching for mutations of particular sequences in cell clones. Large insertions/deletions or the nucleotide substitutions that happen to disrupt or create recognition sequences of restriction enzymes can be detected by Southern blotting after digestion with the appropriate enzymes (restriction fragment length polymorphism, RFLP) (Botstein *et al.*, 1980). Other changes can be detected by any of several recently developed methods. The denaturing gradient gel electrophoresis technique (Myers *et al.*, 1985) takes advantage of the facts that the mobility of partially melted DNA is virtually zero in polyacrylamide gel electrophoresis and that the site of mismatch tends to melt early. Oligonucleotide hybridization (Conner *et al.*, 1983) detects

base changes within the region covered by the probe, because a duplex of oligonucleotides with even a single mismatch is extremely unstable. Other techniques include the ligase-mediated gene detection technique (Landegren *et al.*, 1988) and RNase digestion of a duplex formed between a probe RNA and the target DNA (Winter *et al.*, 1985).

We have reported a method that detects sequence changes, including single-base substitutions as shifts in electrophoretic mobility (Orita *et al.*, 1989). In this technique, sample DNA is digested with a restriction enzyme, denatured, and subjected to polyacrylamide gel electrophoresis under nondenaturing conditions. The target sequence is then detected after electroblotting and hybridization to a probe. Under nondenaturing conditions, single-stranded DNA has a folded conformation that is stabilized by intrastrand interactions. Consequently, the conformation, and therefore the mobility, is dependent on the sequence. As the mobility shift in our method presumably detects conformational changes caused by sequence alterations, we named the technique single-strand conformation polymorphism (SSCP) analysis.

Here we show a simple and rapid method for detection of most sequence changes, including single nucleotide substitutions. In this method, sequences to be examined are amplified and labeled by the polymerase chain reaction (PCR) (Saiki *et al.*, 1988) using labeled primers or a labeled nucleotide, followed by denaturation and electrophoresis for SSCP analysis (PCR-SSCP analysis). Using this technique, we identified new allelic polymorphisms in *Alu* repeats at several chromosomal loci.

MATERIALS AND METHODS

Cell Lines

Lung carcinoma cell line A549, fibrosarcoma cell line HT1080, and promyelocytic leukemia cell line HL60 were obtained from the Japanese Cancer Research Re-

To whom all correspondence should be addressed.

sources Bank. Colon carcinoma cell line SW480 was from the American Type Culture Collection.

DNA Isolation

DNA was prepared from leukocytes, placentas, or cultured cells by the method of Blin and Stafford (1976).

PCR-SSCP Analysis

Oligonucleotides were synthesized by the phosphoramidite method with a 380A DNA synthesizer and purified with OPC columns (Applied Biosystems). The names and the sequences of the primers used here are: NA61, GGTGAAACCTGTTTGTGGA; NA62, ATACACAGGAAGCCTTCG; KA12, GGCCTGCTGAAAATGACTGA; KA13, GTCCTGCACCAGTAATATGC; F41, CCACATGGAGTCTTCATAAT; F42, CCCAGGAGTACTTATTTTA; ABG3, AAGTTGATGCTGGATAGAGG; ABG4, ATTCTCTTGAGACTACATTG; ADE7, CTGGCAAGTGAACAGGTACA; ADE8, CTCTGCATCAGAGAGGGACA; AAG1, CCTCATTCCTATTAGGGAG; and AAG2, CCTGAGGCACATTAAGACAT.

The 5'-ends of primers (100 pmol) were labeled with [γ - 32 P]ATP (50 pmol, 7000 Ci/mmol, ICN) and polynucleotide kinase (5 U, Boehringer-Mannheim) in 10 μ l of 50 mM Tris-HCl, pH 8.3, 10 mM MgCl₂, 5 mM DTT at 37°C for 30 min. The PCR mixture contained 10 pmol each of the labeled primers (kination products added without purification), 2 nmol each of the four deoxynucleotides, 0.1 μ g of genomic sample DNA, and 0.25 U of *Taq* polymerase in 10 μ l of the buffer specified in the GeneAmp kit (Perkin-Elmer Cetus). In some cases, 1 μ l of [α - 32 P]dCTP (3000 Ci/mmol, 10 mCi/ml, Amersham) and primers without the label were used for the amplification. Thirty cycles of the reaction at 94, 55, and 72°C for 0.5, 0.5, and 1 min, respectively, were run in a Thermocycler (Perkin-Elmer Cetus). A portion of the reaction mixture (1 μ l) was withdrawn and mixed with 100 μ l of 0.1% NaDodSO₄ and 10 mM EDTA. Then 2 μ l of this solution was mixed with 2 μ l of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol, heated at 80°C, and applied (1 μ l/lane) to a 5 or 6% polyacrylamide gel (20 \times 40 \times 0.03 cm, 0.5 cm per lane) containing 90 mM Tris-borate, pH 8.3, 4 mM EDTA, and 10% glycerol when specified. Electrophoresis was performed at 30 V for 1-6 h with cooling using a fan. We used a hair tier (with heater turned off) and an aluminum plate attached to one side of the glass plates for efficient and even cooling. The gel was dried on filter paper and exposed to X-ray film at -80°C for 0.5-12 h with an intensifying screen.

Direct DNA Sequencing

Nucleotide sequences were determined using the asymmetric PCR method (Gyllenstein and Erlich, 1988) with slight modifications. An unequal molar ratio (10 to 1) of the primers was used in 50 cycles of the PCR. The amplification mixture (100 μ l) was diluted with 2 ml of water and concentrated to the original volume using a Centricon 30 microconcentrator (Amicon). After annealing to a 5'-labeled primer was performed, the sequencing reactions were carried out using the termination mixtures of a Sequenase kit (USB) and analyzed on a 6% polyacrylamide gel containing 7 M urea.

RESULTS

PCR-SSCP Analysis of Point Mutations

The PCR technique can efficiently amplify DNA segments of known sequences starting from total genomic DNA (Saiki *et al.*, 1988). The amplified products are usually electrophoresed in gels and detected either by staining or by hybridization after blotting. We found that the products can be detected rapidly when labeled primers are included in the chain reaction. Electrophoresis of the reaction mixture in a thin polyacrylamide gel and direct autoradiography of the gel enabled the detection of the amplified sequence in a short time (Hayashi *et al.*, 1989). We next asked whether the labeled amplified DNA fragments could be analyzed for their possible sequence changes by the SSCP method that we reported previously (Orita *et al.*, 1989).

The first example is the detection of activation by point mutations in *NRAS* of HL60 (Boss *et al.*, 1984) and HT1080 (Brown *et al.*, 1984) cells. Both cell lines carry mutations in codon 61 (HL60, CAA to CTA; HT1080, CAA to AAA). Other than at these positions their sequences are identical to those of the normal gene (data not shown). Primers NA61 and/or NA62 (see Materials and Methods for their sequences) were labeled at the 5'-end and included in the mixture for the PCR to amplify a 103-bp fragment that encompasses the codon. The products were denatured and applied to a nondenaturing polyacrylamide gel. Labeling of one (Figs. 1B and 1C) or both (Fig. 1A) of the primers facilitated identification of the bands that were complementary to each other. In HT1080 DNA, four bands could be detected (Fig. 1A, lane 1) when the PCR was carried out in the presence of two labeled primers. The mobilities of two of the four bands were identical to those of the complementary strands of the fragment from the normal sample (Fig. 1, lane 2). These results indicated that the DNA from HT1080 had two different alleles of the *NRAS* gene, one normal and the other mutated. Figure 1 also shows that DNA from HL60 cells contained only the mutated *NRAS* allele

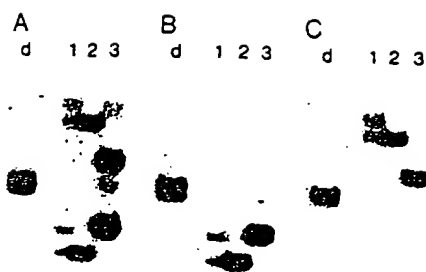


FIG. 1. PCR-SSCP analysis of point mutations in exon 2 of *NRAS*. Both (A) or either (B and C) of the primers flanking exon 2 of *NRAS* were labeled at the 5'-end with ^{32}P and included in the PCR containing 0.1 μg of genomic DNA from HT1080 cells (lane 1), normal leukocytes (lanes 2 and d), and HL60 cells (lane 3). The amplification products were denatured, applied to a 6% polyacrylamide gel (5 mm per lane), and electrophoresed at 4°C. In lanes d, DNA was loaded without denaturation. The exposure for autoradiography was 3 h at -80°C with an intensifying screen.

(lane 3). Furthermore, the mobilities of the single-stranded fragments from the HL60 cells were different from those of the mutated allele of HT1080 cells (lanes 3 vs 1). This implies that the two different mutations can be distinguished by PCR-SSCP analysis.

Next we examined whether point mutations of exon 1 in the *KRAS2* gene could be detected by PCR-SSCP analysis. The lung carcinoma cell line A549 (Valenzuela and Groffen, 1986) and the colon carcinoma cell line SW480 (Capon *et al.*, 1983) have been reported to contain mutations in codon 12 (A549, GGT to AGT; SW480, GGT to GTT). In this experiment we labeled both primers, which spanned a 162-bp segment containing the codon. Figure 2 shows that the mobilities of the separated strands of A549 (lane 1) and SW480 (lane 2) DNAs can be distinguished from those of the normal sample (lane 3). Moreover, none of the cells of the two lines had the normal allele. These results were confirmed by direct DNA sequencing of exon 1 of the *KRAS2* gene in these cells (data not shown).

Conditions That Affect Mobility Shifts

The conformation of single-stranded nucleic acid is presumably determined by the balance between thermal fluctuation and weak local stabilizing forces such as short intrastrand base pairings and base stackings. Therefore, changes in environmental conditions such as temperature and the presence of denaturant are likely to cause a change in conformation, which can be detected in SSCP analysis as an alteration in mobility. That this is indeed true is demonstrated in Fig. 2, where the temperature (A vs B) and the presence of glycerol (A vs C) during electrophoresis are shown to affect the mobilities of separated strands. Yet under all conditions, shifts in mobility by mutations are evident. Also,

the pattern of separations is perfectly reproducible under each condition. In rare cases (6 sequences of approximately 80 examined; this paper and our unpublished data), we observed minor faint bands in the lanes where denatured samples were loaded but not in the lanes where the samples were electrophoresed without denaturation (e.g., Figs. 2A and 2B). The relative intensities and the mobilities of these bands varied depending on the conditions of electrophoresis. We believe that these bands are different conformers of the same sequence and that more than one metastable conformation is sometimes allowed in a given environmental condition.

PCR-SSCP Analysis of DNA Polymorphisms

Given the results described above, we applied PCR-SSCP analysis to the detection of genetic polymorphisms. As a model system, we studied the *D13S2* locus which is on chromosome 13 and closely linked to the *RB* gene (Cavenee *et al.*, 1984). We have shown that several *Hae*III fragments of the *D13S2* locus contain polymorphisms (Orita *et al.*, 1989). One fragment of 430 bp was subcloned from pQD11, which carried a 2.3-kb *Hind*III fragment of the *D13S2* locus (Cavenee *et al.*, 1984), and its nucleotide sequence was determined. Primers corresponding to the sequences at both ends were synthesized (F41 and F42, see Materials and Methods for the sequences) and used for PCR-SSCP analysis of four individuals. The results (Fig. 3A) show that one individual (lane 1 of panels II and III) was heterozygous for this polymorphism. We determined the nucleotide sequences of this fragment from samples 1 and 2 by the asymmetric PCR and direct sequencing

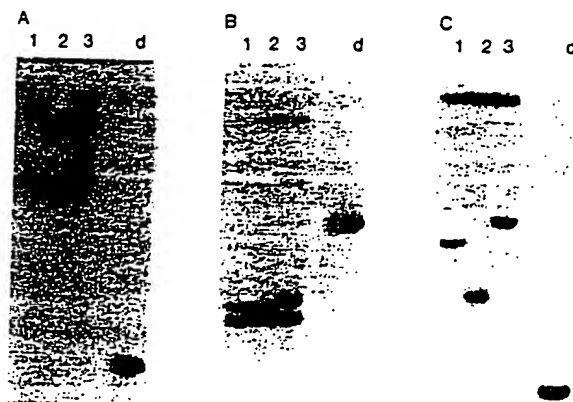


FIG. 2. PCR-SSCP analysis of point mutations in exon 1 of *KRAS*. Genomic DNA from A549 cells (lane 1), SW480 cells (lane 2), and normal leukocytes (lanes 3 and d) was subjected to PCR-SSCP analysis using a pair of labeled primers flanking exon 1 of *KRAS2*, as described in the legend to Fig. 2. In lanes d, DNA was loaded without denaturation. Electrophoresis was carried out in gels (5 mm per lane) containing 10% glycerol at room temperature (A), containing 10% glycerol at 4°C (B), or without glycerol at 4°C (C).



FIG. 3. PCR-SSCP analysis of genetic polymorphism at the *D13S2* locus. DNAs from leukocytes of four unrelated individuals were subjected to PCR-SSCP analysis (A, lanes 1 through 4) as described in the legend to Fig. 2. Electrophoresis was carried out in a 5% gel (5 mm per lane) containing 10% glycerol and run at room temperature. The labeled primers used were (I) F41, (II) F42, and (III) both F41 and F42. See text for the amplified region. The region analyzed in (A) was directly sequenced for the samples 1 and 2 (B). The arrow indicates the band corresponding to a nucleotide substituted in sample 1.

method. The sequence ladder revealed a single nucleotide substitution in sample 1 at base 166 from the 3'-end of the F42 primer (Fig. 3B, arrow). This substitution was confirmed by sequencing the complementary strand (data not shown).

SSCP Analysis of *Alu* Repeats

Alu repeats are present at 10^5 or more copies per human haploid (Deininger *et al.*, 1981). These repeats are believed to have spread throughout the genome by a reverse transcriptase-mediated process (for reviews, see Britten *et al.*, 1988). Most *Alu* repeats seem to have no functional role, since the corresponding sites of nonsimian genomes lack these sequences. We reasoned that such sequences should be rich in polymorphisms because they are under no apparent selective pressure. Using primers that bracket the repeated sequence, PCR-SSCP can be used to search for polymorphisms within such regions. Primers that have sequences in the single-copy regions adjacent to *Alu* repeats at the loci of genes for adenosine deaminase (Wiginton *et al.*, 1988), angiogenin (Kurachi *et al.*, 1985), and β -globin (Henthorn *et al.*, 1986) were synthesized and PCR-SSCP analysis was performed using [α - 32 P]dCTP as a labeled precursor (Fig. 4). Examination of nine unrelated individuals revealed that these segments were highly polymorphic. Heterozygosity was detected for four, five, and four of nine individuals at the loci of adenosine deaminase, angiogenin, and β -globin genes, respectively. The direct sequencing confirmed base substitutions in these alleles and that they were inherited according to Mendel's law (data not shown).

DISCUSSION

The method of detecting nucleotide sequence polymorphisms described here is simple, fast, and efficient. The target sequence is amplified and labeled simultaneously by the PCR using labeled primers or labeled deoxynucleotide. Therefore, neither restriction enzyme digestion nor hybridization is required.

As primers, we have chosen sequences with balanced composition in the desired regions. Using these primers

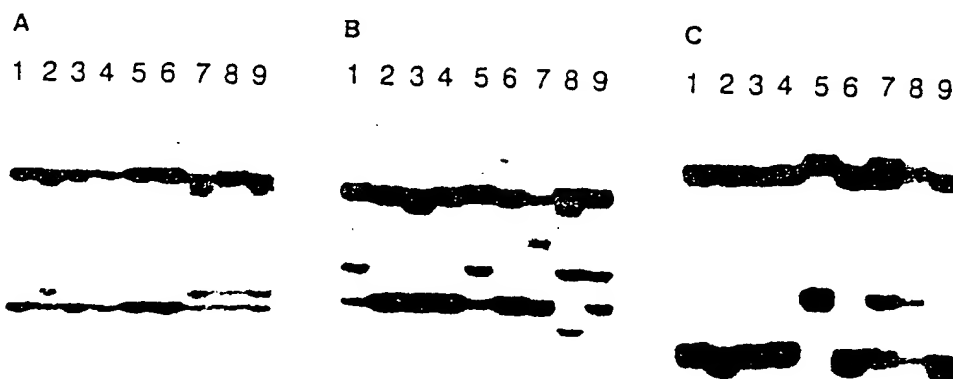


FIG. 4. PCR-SSCP analysis of the *Alu* repeats. DNA samples of unrelated individuals were analyzed using pairs of labeled primers that encompassed the *Alu* repeats in loci of genes for adenosine deaminase (A), angiogenin (B) and β -globin (C). Primers and lengths of the amplified regions were (A) ADE7 and ADE8, 381 bp; (B) AAG1 and AAG2, 361 bp; and (C) ABG3 and ABG4, 417 bp. Electrophoresis was carried out in a 5% polyacrylamide gel (5 mm per lane) containing 10% glycerol and run at room temperature.

under the PCR conditions described above, no spurious reaction products were observed, except in a few exceptional cases. One such case was when primers were located in exon sequences of *RAS* genes (data not shown). Perhaps it is advisable to use nonexon sequences as primers to avoid possible unwanted amplifications of several sequences which might be shared among members of the gene families or pseudogenes. Another exception occurred when a region within a cluster of *Alu* repeats in the β -tubulin locus was examined (data not shown). In this case, some of the runthrough products in the first cycle that ended within the repeated sequence may have served as primers of the following cycles, and initiated cycles of artifactual amplifications.

The effect of sequence change on electrophoretic mobility is unpredictable. It is true that some of the sequence change may not appreciably affect the mobility. However, we observed mobility shift with SSCP analysis (this report and our unpublished data) in all 12 arbitrarily chosen tumor cell lines that are known to contain mutated *HRAS*, *KRAS*, or *NRAS*.

In a typical PCR, up to 10% of the substrates (primers and deoxynucleotides) are incorporated into the amplified product. Thus, the efficiency of labeling of target sequences in PCR using labeled primers or labeled nucleotide is extremely high when compared to the efficiency in Southern blotting experiments in which much smaller portions of the label attach to the target sequences by hybridization. Consequently, in the current method, the time of exposure to X-ray film is much shorter than that in the RFLP experiments and the entire procedure including exposure time can be completed within 24 h.

The high radioactivity in the target sequence also permitted the use of a thin polyacrylamide gel (Sanger and Coulson, 1977). With a 0.3-mm gel, a steep voltage gradient can be applied without serious Ohmic heating, so that the time required for electrophoresis is shortened. This improvement is particularly important since the conformation of single-stranded DNA is sensitive to temperature. Also, the high resolution of a thin gel is obviously advantageous for detection of subtle conformational changes in the samples.

The electrophoretic mobility of single-stranded DNA in polyacrylamide gels is strongly dependent on environmental conditions, as can be seen in Fig. 2. We examined several different conditions of electrophoresis in a search for better resolution in the PCR-SSCP analysis. In the analyses of 12 sequences that contain different mutated *RAS* and several *ALU* repeats of different loci, we performed electrophoresis at room temperature or at 4°C in the presence or absence of 10% glycerol. In our experience, mobility shifts caused by base substitutions of most sequences were best resolved

when electrophoresis was carried out at room temperature in the presence of 10% glycerol.

PCR-SSCP analysis requires prior information on the representative sequence for the design of primers. Within this limitation, this method may be of choice when a large number of samples must be examined, such as in linkage analysis of known sequence, survey of activated oncogenes in cancer tissues, or prenatal diagnosis for the presence or absence of a particular allele.

The human genome is believed to contain, on average, one polymorphism every few hundred base pairs (Botstein *et al.*, 1980). Therefore, PCR-SSCP analysis of randomly chosen regions of several hundred base pairs is likely to reveal sequence polymorphisms, especially when the regions are apparently nonfunctional. For example, we examined *Alu* repeat sequences using single-copy sequences bracketing the repeats as primers, and found that such interspersed repeated sequences can be examined by PCR-SSCP analysis. With other methods that involve hybridization to probes, analysis of polymorphisms within such sequences is difficult. As expected, we found that three arbitrarily chosen *Alu* repeats (in loci of genes for adenosine deaminase, angiogenin, and β -globin) were highly polymorphic. Because of their abundance, the search and linkage analyses of polymorphisms in the *Alu* repeats in the vicinity of genes or even in the anonymous sequences using PCR-SSCP may be useful in the construction of a linkage map of the total human genome.

ACKNOWLEDGMENTS

This work was supported in part by a grant-in-aid from the Ministry of Health and Welfare for a Comprehensive 10-Year Strategy for Cancer Control, Japan, and by grants-in-aid from the Ministry of Science, Education and Culture, Japan, and from the Special Coordination Fund of the Science and Technology Agency of Japan. M. Orita and Y. Suzuki were recipients of Research Resident Fellowships from the Foundation for Promotion of Cancer Research.

REFERENCES

1. BLIN, N., AND STAFFORD, D. M. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* 3: 2303-2308.
2. BOSS, J. L., DE VRIES, V., JANSSEN, A. M., VEENEMAN, G. H., VAN BOOM, J. H., AND VANDER EB, A. J. (1984). Three different mutations in codon 61 of the human *N-ras* gene detected by synthetic oligonucleotide hybridization. *Nucleic Acids Res.* 12: 9155-9163.
3. BOTSTEIN, D., WHITE, R. L., SKOLNICK, M., AND DAVIS, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* 69: 201-205.
4. BRITTEN, R. J., BARON, W. F., STOUT, D. B., AND DAVIDSON, E. H. (1988). Source and evolution of human *Alu* repeated sequences. *Proc. Natl. Acad. Sci. USA* 85: 4770-4774.

5. BROWN, R., MARSHAL, C. J., PENNIE, S. G., AND HALL, A. (1984). Mechanism of activation of an *N-ras* gene in the human fibrosarcoma cell line HT1080. *EMBO J.* 3: 1321-1326.
6. CAPON, D. J., SEEBERG, P. H., MCGRATH, J. P., HAYFLICK, J. S., EDMAN, U., LEVINSON, A. D., AND GOEDDEL, D. V. (1983). Activation of *Ki-ras* 2 gene in human colon and lung carcinomas by two different point mutations. *Nature (London)* 304: 507-513.
7. CAVENEE, W., LEACH, R., MAHANDAS, T., PEARSON, P., and WHITE, R. (1984). Isolation and regional localization of DNA segments revealing polymorphic loci from human chromosome 13. *Amer. J. Hum. Genet.* 36: 10-24.
8. CONNER, B. J., REYES, A. A., MORIN, C., ITAKURA, K., TEPLITZ, R. L., AND WALLACE, R. B. (1983). Detection of sickle cell beta-globin allele by hybridization with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* 80: 278-282.
9. DEININGER, P. L., JOLLY, D. J., RUBIN, C. M., FRIEDMANN, T., AND SCHMIDT, C. W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* 151: 17-33.
10. GYLLENSTEN, U. B., AND ERLICH, H. A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. *Proc. Natl. Acad. Sci. USA* 85: 7652-7656.
11. HAYASHI, K., ORITA, M., SUZUKI, Y., AND SEKIYA, T. (1989). Use of labeled primers in polymerase chain reaction (LP-PCR) for a rapid detection of the product. *Nucleic Acids Res.* 17: 3605.
12. HENTHORN, P. S., MAGER, D. L., HUISMAN, H. J., AND SMITHIES, O. (1986). A gene deletion ending within a complex array of repeated sequences 3' to the human beta-globin gene cluster. *Proc. Natl. Acad. Sci. USA* 83: 5194-5198.
13. KURACHI, K., DAVIE, E. W., STRYDOM, D. J., RIORDAN, J. F., AND VALLEE, B. L. (1985). Sequence of the cDNA for angiogenin, a human angiogenesis factor. *Biochemistry* 24: 5494-5499.
14. LANDEGREN, U., KAISER, R., SANDERS, J., AND HOOD, L. (1988). A ligase-mediated gene detection technique. *Science* 241: 1077-1080.
15. MYERS, R. M., LUMELSKY, N., LERMAN, L., AND MANIATIS, T. (1985). Detection of single base substitutions in total genomic DNA. *Nature (London)* 313: 495-498.
16. ORITA, M., IWAHANA, H., KANAZAWA, H., HAYASHI, K., AND SEKIYA, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86: 2766-2770.
17. SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B., AND ERLICH, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487-491.
18. SANGER, F., AND COULSON, A. R. (1977). The use of thin acrylamide gels for DNA sequencing. *FEBS Lett.* 87: 107-110.
19. VALENZUELA, D. M., AND GROFFEN, J. (1986). Four human carcinoma cell lines with novel mutations in position 12 of c-K-ras oncogene. *Nucleic Acids Res.* 14: 843-852.
20. WIGINTON, D. A., KAPLAN, D. J., STATES, J. C., AKESON, A. L., PERME, C. M., BILYK, I. J., VAUGHN, A. J., LATTIER, D. L., AND HUTTON, J. J. (1988). Complete sequence and structure of the gene for human adenosine deaminase. *Biochemistry* 25: 8234-8244.
21. WINTER, E., YAMAMOTO, F., ALMOGUERA, C., AND PERUCHO, M. (1985). A method to detect and characterize point mutations in transcribed genes: amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells. *Proc. Natl. Acad. Sci. USA* 82: 7575-7579.

Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms

(mobility shift of separated strands/point mutation/restriction fragment length polymorphism)

MASATO ORITA, HIROYUKI IWAHANA*, HIROSHI KANAZAWA†, KENSHI HAYASHI, AND TAKAO SEKIYA‡

Oncogene Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan

Communicated by Takashi Sugimura, December 29, 1988

ABSTRACT We developed mobility shift analysis of single-stranded DNAs on neutral polyacrylamide gel electrophoresis to detect DNA polymorphisms. This method follows digestion of genomic DNA with restriction endonucleases, denaturation in alkaline solution, and electrophoresis on a neutral polyacrylamide gel. After transfer to a nylon membrane, the mobility shift due to a nucleotide substitution of a single-stranded DNA fragment could be detected by hybridization with a nick-translated DNA fragment or more clearly with RNA copies synthesized on each strand of the DNA fragment as probes. As the mobility shift caused by nucleotide substitutions might be due to a conformational change of single-stranded DNAs, we designate the features of single-stranded DNAs as single-strand conformation polymorphisms (SSCPs). Like restriction fragment length polymorphisms (RFLPs), SSCP s were found to be allelic variants of true Mendelian traits, and therefore they should be useful genetic markers. Moreover, SSCP analysis has the advantage over RFLP analysis that it can detect DNA polymorphisms and point mutations at a variety of positions in DNA fragments. Since DNA polymorphisms have been estimated to occur every few hundred nucleotides in the human genome, SSCP s may provide many genetic markers.

The nucleotide sequences of DNAs in humans are not identical in different individuals. Nucleotide substitutions have been estimated to occur every few hundred base pairs in the human genome (1). Nucleotide sequence polymorphism has been detected as restriction fragment length polymorphism (RFLP). RFLP analysis of family members has been used to construct a genetic linkage map of the human genome (2, 3), and this analysis has also revealed the chromosomal locations of genetic elements involved in hereditary diseases such as Huntington disease (4), adult polycystic kidney disease (5), cystic fibrosis (6-8), Alzheimer disease (9, 10), and Duchenne muscular dystrophy (11, 12). Thus prenatal diagnosis of diseases such as cystic fibrosis is possible with RFLP probes. Recently, RFLP analysis has indicated specific loss of heterozygosity at particular loci on chromosomes in cancerous portions of tissues in several human cancers, including retinoblastoma, Wilms tumor, small cell carcinoma of the lung, renal cell carcinoma, bladder carcinoma, breast carcinoma, meningioma, acoustic neuroma (see ref. 13 for a review), colorectal carcinoma (14, 15), and multiple endocrine neoplasia type 1- or type 2-associated carcinomas (16, 17). This loss of heterozygosity suggests the involvement of recessive mutation of particular genes in development of these cancers.

Although RFLP s are very useful for distinguishing two alleles at chromosomal loci, they can be detected only when DNA polymorphisms are present in the recognition se-

quences for the corresponding restriction endonucleases or when deletion or insertion of a short sequence is present in the region detected by a particular probe. To identify DNA polymorphisms more efficiently, Noll and Collins used a simplified method of denaturing gradient gel electrophoresis (18) that had been developed by Myers *et al.* (19). As analysis of mobility shift [probably due to a conformational change of single-stranded DNAs on polyacrylamide gel electrophoresis (20)] has been used to detect point mutations (21), in this work we examined whether the mobility shift of single-stranded DNA caused by a single nucleotide substitution could be used to detect nucleotide sequence polymorphisms. The results indicated that mobility shift analysis is an efficient method for detecting DNA polymorphisms and for distinguishing the two alleles at chromosomal loci.

MATERIALS AND METHODS

Cell Lines. The human bladder carcinoma cell line T24 was obtained from the American Type Culture Collection. The human malignant melanoma cell line SK2 was established from a tissue that had been maintained in *nude* mice (22).

DNA Isolation. High molecular weight DNA was prepared from human leukocytes or cultured human tumor cell lines by the method of Blin and Stafford (23).

Plasmids. Plasmid pNCO106 was prepared by inserting a 2.9-kilobase pair (kb) *Sac* I fragment of the *HRAS1* gene from SK2 cells into pUC19 (24). Plasmid pT22 was constructed by inserting a 6.6-kb *Bam* HI fragment of the *HRAS1* gene from T24 cells into pBR322 (a gift from M. Wiglar, Cold Spring Harbor Laboratory).

Subcloning and Sequencing of DNA Fragments. From pNCO106 and pT22, a 371-base-pair (bp) *Pst* I fragment carrying exon 1 and a 298-bp *Pst* I fragment containing exon 2 of the *HRAS1* gene were isolated and subcloned into the pGEM-2 vector (Promega Biotec). The nucleotide sequences of the subcloned fragments were determined by the dideoxynucleotide method (25), using Sequenase (United States Biochemical) and the SP6 or T7 promoter primer (Promega Biotec).

Analysis of Single-Strand Conformation Polymorphisms (SSCPs). High molecular weight DNA (20 μ g) was digested completely with restriction endonucleases under the conditions recommended by the suppliers. The reaction mixture was extracted once with phenol/chloroform (1:1, vol/vol) and once with chloroform. After addition of 0.1 vol of 3 M sodium acetate, DNA fragments were precipitated from the

Abbreviations: RFLP, restriction fragment length polymorphism; SSCP, single-strand conformation polymorphism.

*Present address: Department of Internal Medicine, School of Medicine, The University of Tokushima, 3-18-15 Kuramoto-cho, Tokushima 770, Japan.

†Present address: Department of Applied Biology, Faculty of Technology, Okayama University, 3-1-1 Tsushima-cho, Okayama 700, Japan.

‡To whom reprint requests should be addressed.

aqueous phase by addition of 2.5 vol of ethanol. Strands were separated out by the method of Maxam and Gilbert (20) with a slight modification. DNA precipitates were dissolved in 20 μ l of denaturing solution (0.3 M NaOH/1 mM EDTA) and then mixed with 3 μ l of 50% (vol/vol) glycerol/0.25% xylene cyanol/0.25% bromophenol blue. The mixture was applied to a neutral 5% polyacrylamide gel (20 \times 40 \times 0.2 cm) with or without 10% glycerol in a well of 10 mm width and subjected to electrophoresis in 90 mM Tris-borate, pH 8.3/4 mM EDTA at 180 V for 12–36 hr at 17°C. DNA fragments in the gel were then transferred to a nylon membrane (Hybond-N, Amersham) by electrophoretic blotting in 0.025 M sodium phosphate, pH 6.5, at 1 A for 2 hr at 4°C by the procedure recommended by the membrane supplier. The membrane was then dried and baked at 80°C for 2 hr. Hybridization with 32 P-labeled DNA probes was performed in 50% (vol/vol) formamide/6 \times SSC (1 \times SSC is 0.15 M sodium chloride/0.015 M sodium citrate, pH 7.0)/10 mM EDTA/5 \times Denhardt's solution (1 \times Denhardt's solution is 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone)/0.5% NaDodSO₄ containing denatured salmon sperm DNA at 100 μ g/ml and 10% dextran sulfate at 42°C for 16 hr. The blots were washed twice in 2 \times SSC/0.1% NaDodSO₄ for 30 min at 65°C and then once in 0.1 \times SSC at 65°C for 10 min. Autoradiography was carried out at -80°C for 2–7 days by exposing the membranes to x-ray film (XAR-5, Kodak) with an intensifying screen (Cronex Lightning Plus, DuPont).

Analysis of RFLP. RFLP analysis was performed as described (26). High molecular weight DNA (5 μ g) was digested with an appropriate restriction endonuclease and the digest was fractionated by electrophoresis in a 0.7% agarose gel.

DNA Probes for Hybridization. Cloned *Pst* I fragments 371 and 298 bp long carrying exon 1 and 2 of the normal human *HRAS* gene (27), respectively, were used as specific probes for the corresponding exons. The 2.8-kb *Hind*III fragment isolated from phage 9D11 (28), provided by the Japanese Cancer Research Resources Bank, was used as a specific probe for the *D13S2* locus on human chromosome 13 (29). Probes were labeled to a specific activity of 2–10 $\times 10^8$ cpm/ μ g by nick-translation (30) with [α - 32 P]dCTP (3000 Ci/mmol; 1 Ci = 37 GBq) as a radioactive substrate.

RNA Probes for Hybridization. Single-stranded RNA probes were prepared by the method of Melton *et al.* (31) with plasmid constructs carrying the fragments used as DNA probes in the pGEM-2 vector as templates. RNA synthesis on each strand of the templates was carried out with T7 RNA polymerase (TOYOBO, Tokyo) or SP6 RNA polymerase (Amersham) in the presence of [α - 32 P]UTP as a radioactive substrate. Concentration of UTP was adjusted to 500 μ M by adding the nonradioactive nucleotide (final specific activity, 40 Ci/mmol) to ensure synthesis of full-length RNA copies. The hybridization conditions and washing procedures were the same as those for DNA probes.

RESULTS

Mobility Shift by Single Base Substitution. To determine whether a single base substitution altered the mobility of single-stranded DNAs on neutral polyacrylamide gel electrophoresis, we separated *Pst* I fragments carrying exon 1 or 2 of the human *HRAS* gene, whose nucleotide sequences are known. In the human melanoma cell line SK2, one of the two alleles of the *HRAS* gene is known to be activated by point mutation at codon 61 in exon 2 (32) and also amplified about 0-fold (33). The human bladder carcinoma cell line T24 has been reported to contain only one allele of the *HRAS* gene, which carries a mutated codon 12 in exon 1 (34, 35). From plasmid constructs pNCO106 and pT22, containing the transforming allele of the *HRAS* gene of SK2 and T24 cells, respectively, a 371-bp *Pst* I fragment carrying exon 1 of the

gene was isolated and subcloned in the pGEM-2 vector. Similarly, a 298-bp *Pst* I fragment carrying exon 2 of the *HRAS* gene was isolated from the same plasmid constructs and subcloned. By determination of the total nucleotide sequences of the subcloned fragments, we confirmed the single nucleotide substitution at codon 12 in the 371 nucleotides of the *Pst* I fragment between the SK2 gene and the T24 gene (GGC in the SK2 gene and GTC in the T24 gene). The nucleotide sequences of the 298-bp *Pst* I fragments carrying exon 2 of the SK2 and T24 genes were also confirmed to differ from each other by only one nucleotide in codon 61 (CTG in the SK2 gene and CAG in the T24 gene). After denaturation in alkaline solution, these cloned *Pst* I fragments were subjected to electrophoresis in neutral 5% polyacrylamide gel. The separated strands were then transferred to a nylon membrane by electrophoretic blotting and hybridized with 32 P-labeled DNA probes. As shown in Fig. 1A, the pair of separated strands of the *Pst* I fragment carrying exon 1 of the T24 gene (lane 2) moved slightly faster than those of the SK2 gene (lane 1). In the case of the *Pst* I fragment carrying exon 2, the mobilities of the separated strands of the SK2 gene (Fig. 1A, lane 3) were significantly different from those of the T24 gene (lane 4). Three bands were observed in the sample from the SK2 gene. Hybridization with single-stranded RNA probes showed that the bands with the fastest and the slowest mobilities were from the same strand of the fragment, while the middle band corresponded to the complementary strand (data not shown). Usually the slowest-moving band was the major one from the particular strand and the ratio of the slowest and the fastest bands varied depending on the conditions of electrophoresis, especially the temperature of the running gels. These results suggested that a particular single-stranded DNA could take at least two different molecular shapes, depending on the conditions of electrophoresis.

In the system containing homogeneous cloned DNA fragments, we could demonstrate mobility shift of single-stranded DNAs due to a single base substitution. To determine whether the same mobility shift could be observed in the presence of DNA fragments other than a target fragment, we digested genomic DNAs from the two tumor cell lines SK2 and T24 with *Pst* I and subjected the total digests to electrophoresis in neutral polyacrylamide gel after denaturation. As shown in Fig. 1B, the patterns of the separated strands of the fragments carrying exon 1 or 2 of the *HRAS* gene from the genomic DNAs were essentially the same as those of the cloned fragments. This result indicated that the mobility shift due to a single base substitution of a single-stranded DNA fragment in total digests of genomic DNA

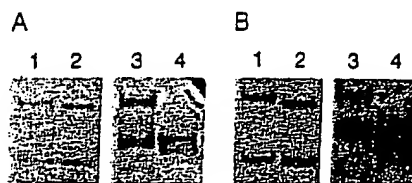


FIG. 1. Mobility shift of single-stranded DNA fragments due to a single base substitution. (A) Plasmid clones (2 pg) of fragments carrying exon 1 (371 bp) and exon 2 (298 bp) of the *HRAS* gene from malignant melanoma SK2 cells (lanes 1 and 3, respectively) and from bladder carcinoma T24 cells (lanes 2 and 4, respectively) were digested with *Pst* I. (B) Total genomic DNAs (20 μ g) from SK2 cells (lanes 1 and 3) and from T24 cells (lanes 2 and 4) were digested with *Pst* I. After denaturation, the fragments produced were subjected to electrophoresis in neutral polyacrylamide gel without glycerol. Single-stranded DNAs were transferred to a nylon membrane and hybridized with the 32 P-labeled DNA probe for exon 1 of the *HRAS* gene (lanes 1 and 2 in A and B) and the probe for exon 2 of the gene (lanes 3 and 4 in A and B).

could be detected and was not influenced by the presence of a large amount of unrelated DNA fragments.

SSCP Analysis of Human DNA at the *D13S2* Locus. The above results encouraged us to apply the mobility shift of single-stranded DNA due to a single base substitution to detection of nucleotide sequence polymorphisms of a particular fragment and, as can be done with RFLPs, to distinguishing two alleles at chromosomal loci. As the mobility shift might be due to a conformational change of the single-stranded DNAs, we designated the polymorphisms detected by the method as SSCPs.

Leukocyte DNA samples from 19 individuals (10 unrelated and 9 in two families) were digested with *Hae* III, and SSCPs of the fragments obtained from a region of about 3 kb at the *D13S2* locus on chromosome 13 were analyzed. When the digests were subjected to electrophoresis without denaturation and hybridized with the 32 P-labeled 2.8-kb *Hind*III fragment as a specific probe for the *D13S2* locus, five distinct double-stranded DNA fragments (F1 to F5 in order of size) without any RFLP were observed in all DNA samples. The results on DNA samples 1 and 2 are shown in Fig. 2A as examples. In contrast with the double-stranded fragments, separated strands of the same DNA fragments showed SSCPs with considerable frequency. Representative results are shown in Fig. 2B–D. When nick-translated DNA was used as a probe, SSCPs were apparently observed in at least one of the four fragments (F2 to F5) in all four DNA samples (Fig. 2B). The mobility shift of one of the strands of fragment F4 in sample 1 was especially marked. However, the mobility shifts of single strands in other fragments were small and therefore the difference of the shifts was not clear when both strands of the fragments were hybridized with the nick-translated probe. To overcome this disadvantage, RNA copies (RNA 1 and 2 in Fig. 2C and D) of each strand of the *D13S2* DNA fragment were prepared separately and used as probes for hybridization. As shown in Fig. 2C and D, with either the RNA 1 or RNA 2 probe SSCPs were clearly detected in all fragments except fragment F1. In Fig. 2E, the alleles distinguished by SSCPs are summarized. SSCPs found in fragment F2 by using the RNA 1 probe could distinguish alleles with three different mobilities, designated as "slow" (s), "fast" (f), and "very fast" (vf). In addition to these three

alleles, the SSCP analysis of the other DNA sample shown in Fig. 3A revealed the presence of an allele with "very slow" (vs) mobility in the fragment. The SSCPs of the other fragments, F3, F4, and F5, could also distinguish at least two alleles with "slow" (s) or "fast" (f) mobility. Analysis of 19 DNA samples revealed that mobility shifts found in F4 and F5 were coincidental.

Mendelian Inheritance of SSCPs. To confirm that the observed SSCPs of the *Hae* III fragments of the region at the *D13S2* locus were due to allelic variants of true Mendelian traits, we analyzed the DNAs of nine individuals in two related families. In Fig. 3A, SSCPs of fragments F2, F3, and F4 and the alleles identified are indicated. In each family, the genotypes of the progenies were consistent with the parental genotypes.

Relationship Between SSCPs and RFLPs. The same 19 DNA samples analyzed for SSCPs were also subjected to RFLP analysis. The DNAs were digested with *Msp* I or *Taq* I and RFLPs were detected by hybridization with the 32 P-labeled DNA probe for the *D13S2* locus. Of the 19 DNA samples digested with *Msp* I, five samples (sample 2 in Fig. 2, data not shown, samples 2, 3, 5, and 8 in Fig. 3B) showed RFLP. By *Taq* I digestion, RFLP was observed in only one of the DNA samples (sample 2 in Fig. 2, data not shown). Therefore, RFLP analysis revealed heterozygosity at the *D13S2* locus in only 5 of 19 individuals, while with SSCP analysis heterozygosity at the locus was found in at least one of the four *Hae* III fragments in 18 of the 19 DNA samples. This fact demonstrates that SSCP analysis is a superior tool for detection of genetic polymorphisms.

Factors Affecting SSCP Analysis. The mobility shift of single-stranded DNAs with DNA polymorphisms observed on neutral polyacrylamide gel electrophoresis is most likely due to conformational variations of the molecules. The conformation of single-stranded nucleic acid is expected to be affected by environmental factors such as the temperature of the gel during electrophoresis, the concentration of electrophoresis buffer, and the presence of denaturing agents in gels. The mobility shift of the *Pst* I fragments carrying exon 1 of the *HRAS1* gene shown in Fig. 1A (lanes 1 and 2) was clearly observed on electrophoresis at 17°C but not prominently at 23°C (data not shown). The pattern of the separated strands

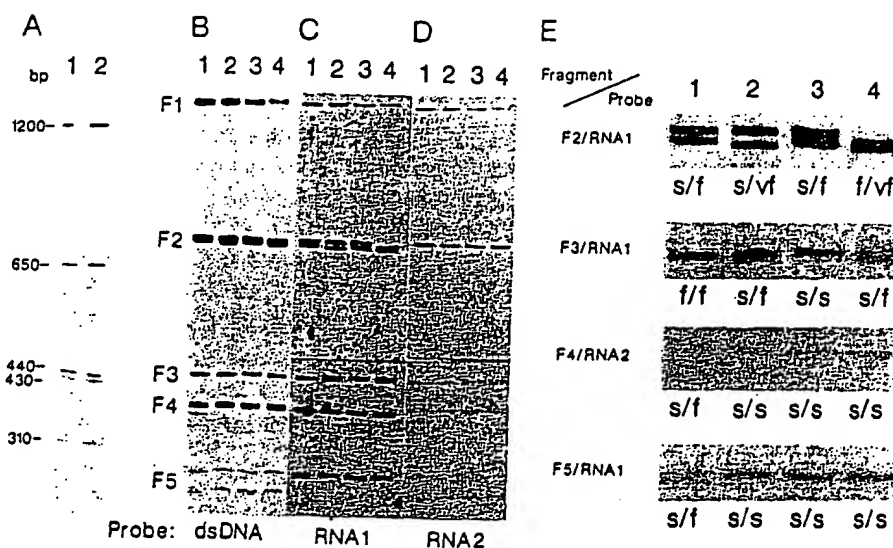


FIG. 2. SSCP analysis of human DNAs at the *D13S2* locus. DNA samples 1–4 were prepared from leukocytes of four unrelated individuals and digested with *Hae* III. The resultant fragments were subjected to electrophoresis in neutral polyacrylamide gel containing 10% glycerol before (A) and after (B, C, and D) denaturation. DNAs in the gel were transferred to a nylon membrane and then hybridized with the 32 P-labeled double-stranded DNA (dsDNA) probe for the *D13S2* locus (A and B) and with the 32 P-labeled single-stranded RNA probes for the *D13S2* locus (RNA1 in C and RNA2 in D). The five fragments produced from the *D13S2* region by *Hae* III digestion were designated as F1 to F5 in order of size. Alleles identified by SSCPs are indicated in E with higher magnifications of informative fragments observed in C or D.

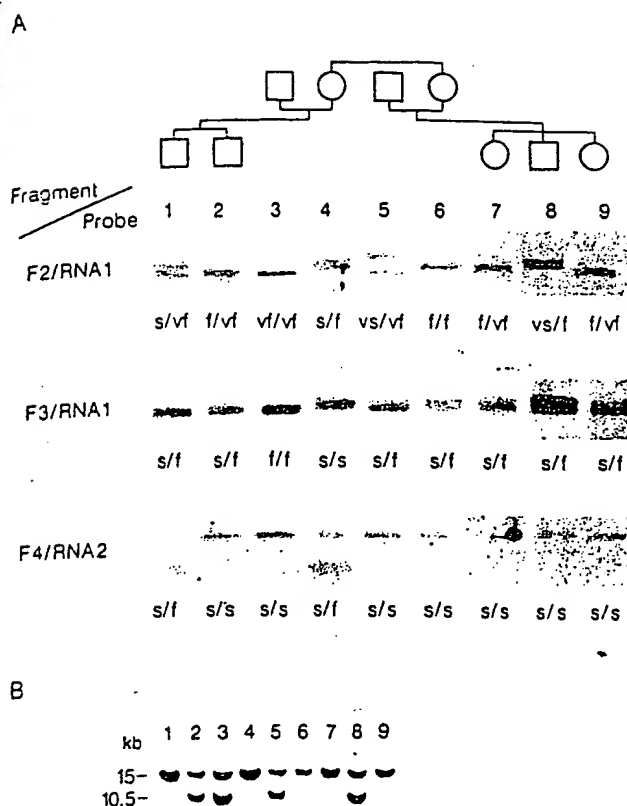


FIG. 3. SSCP and RFLP analyses of family members. (A) Leukocyte DNAs (20 μ g) from the family members indicated at the top (o, females; \square , males) were subjected to SSCP analysis using the *D13S2* probe as described in the legend for Fig. 2. As the mobility shifts found in fragments F4 and F5 were the same, the results with fragment F5 are not shown. (B) The leukocyte DNAs (5 μ g) digested with *Msp* I were subjected to RFLP analysis using 32 P-labeled dsDNA as a probe for the *D13S2* locus.

of the fragments carrying exon 2 of the gene observed at 17°C and shown in Fig. 1A (lanes 3 and 4) was also altered at 23°C. Thus, the higher temperature might destroy some semistable conformations. The concentration of the running buffer also affected the mobility shift. When electrophoresis of the *Pst* I fragments analyzed in Fig. 1A was performed in a buffer of lower concentration (45 mM Tris-borate, pH 8.3/2 mM EDTA) at 17°C, the mobility shifts observed were similar to those at the higher temperature (23°C). Presence of 10% glycerol in gels also affected the mobility shift. However the effect of glycerol was rather complicated and mobility shifts due to DNA polymorphisms were often enhanced by this reagent. For example, the mobility shifts observed in Fig. 2 were enhanced when electrophoresis was performed in gel containing 10% glycerol. On the other hand, the mobility shift shown in Fig. 1 was reduced by the presence of 10% glycerol in the gel.

DISCUSSION

By neutral polyacrylamide gel electrophoresis, we could separate two single-stranded DNA fragments in which the nucleotide sequences differed at only one position. The mobility shift due to a single base substitution could be observed not only in cloned fragments but also in fragments of total genomic DNA after restriction endonuclease digestion. We applied the method to detect nucleotide sequence polymorphisms in human genomic DNA and could observe the mobility shift of single-stranded DNA by using a genomic sequence probe arbitrarily chosen. Single-stranded DNAs of

the same nucleotide length can be separated by polyacrylamide gel electrophoresis, probably due to a difference in their predominant semistable conformations (20). The mobility shift of single-stranded DNAs with DNA polymorphisms observed on gel electrophoresis might also be due to conformational change, and so we designated the features of DNAs as SSCPs. We do not know whether nucleotide substitution at any position in a fragment can be detected by SSCP analysis, but DNA polymorphisms at a variety of positions in a fragment could cause a difference in its conformation and result in change in mobility of the single strands on gel electrophoresis. Therefore, we thought that DNA polymorphism could be detected more frequently by SSCP analysis than by RFLP analysis, and our experimental results revealed that this was in fact the case. Like RFLP analysis, SSCP analysis is simple and does not require complicated instruments or specialized techniques.

As we confirmed that the observed SSCPs were due to allelic variation of true Mendelian traits, SSCP analysis of DNA fragments could be a useful and simple method for elucidating the human genetic linkage map by studies on families. Because DNA polymorphisms have been estimated to occur once every few hundred nucleotides of the human genome (1) and SSCP analysis can reveal nucleotide substitutions at various positions in a fragment, any restriction endonuclease fragment with a nucleotide length suitable for strand separation may provide information for distinguishing two alleles. Therefore, in theory, on a nylon membrane carrying separated strands of all possible fragments of genomic DNA, DNA polymorphisms at any chromosomal locus can be detected by repeated hybridization of the membrane with a variety of probes.

SSCP analysis can also be used to locate genetic elements involved in hereditary diseases and to detect DNA aberrations in human cancers. Comparison of DNA fragments from cancerous portions of tissues with those from normal portions by SSCP analysis can reveal amplified alleles of particular genes and loss of heterozygosity at particular chromosomal loci. A remarkable advantage of SSCP analysis is that it can be used to detect point mutations at various positions in a fragment. Recently, by means of the DNA polymerase chain reaction (PCR), a DNA segment of a single cell or a single sperm has been amplified to an amount sufficient for analysis by hybridization (36). Our preliminary result suggested that SSCP analysis of DNA segments amplified by PCR technique could be useful for diagnosis of genetic aberrations.

This work was supported in part by a grant-in-aid from the Ministry of Health and Welfare for a Comprehensive 10-Year Strategy for Cancer Control, Japan, and a grant from the Special Coordination Fund of the Science and Technology Agency of Japan. M.O. and H.I. were recipients of Research Resident Fellowships from the Foundation for Promotion of Cancer Research.

- Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. & Schmidtke, J. (1985) *Hum. Genet.* 69, 201-205.
- Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* 32, 314-331.
- Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lander, E. S., Botstein, D., Akots, G., Rediker, K. S., Gravius, T., Brown, V. A., Rising, M. B., Parker, C., Powers, J. A., Watt, D. E., Kauffman, E. R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T. R., Ng, S., Schumm, J. W., Braman, J. C., Knowlton, R. G., Barker, D. F., Crooks, S. M., Lincoln, S. E., Daly, M. J. & Abrahamson, J. (1987) *Cell* 51, 319-337.
- Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, I.,

- Bonilla, E. & Martin, J. B. (1983) *Nature (London)* 306, 234-238.
5. Reeders, S. T., Breuning, M. H., Davies, K. E., Nicholls, R. D., Jarman, A. P., Higgs, D. R., Pearson, P. L. & Weatherall, D. J. (1985) *Nature (London)* 317, 542-544.
6. Knowlton, R. G., Cohen-Haguenauer, O., Van Cong, N., Frézal, J., Brown, V. A., Barker, D., Braman, J. C., Schumm, J. W., Tsui, L.-C., Buchwald, M. & Donis-Keller, H. (1985) *Nature (London)* 318, 380-382.
7. White, R., Woodward, S., Leppert, M., O'Connell, P., Hoff, M., Herbst, J., Lalouel, J.-M., Dean, M. & Vande Woude, G. (1985) *Nature (London)* 318, 382-384.
8. Wainwright, B. J., Scambler, P. J., Schmidtke, J., Watson, E. A., Law, H.-Y., Farrall, M., Cooke, H. J., Eiberg, H. & Williamson, R. (1985) *Nature (London)* 318, 384-385.
9. Tanzi, R. E., Gusella, J. F., Watkins, P. C., Bruns, G. A. P., St George-Hyslop, P., Van Keuren, M. L., Patterson, D., Pagán, S., Kurnit, D. M. & Neve, R. L. (1987) *Science* 235, 880-884.
10. St George-Hyslop, P. H., Tanzi, R. E., Polinsky, R. J., Haines, J. L., Nee, L., Watkins, P. C., Myers, R. H., Feldman, R. G., Pollen, D., Drachman, D., Growdon, J., Bruni, A., Foncin, J.-F., Salmon, D., Frommelt, P., Amaducci, L., Sorbi, S., Placentini, S., Stewart, G. D., Hobbs, W. J., Conneally, P. M. & Gusella, J. F. (1987) *Science* 235, 885-890.
11. Monaco, A. P. & Kunkel, L. M. (1987) *Trends Genet.* 3, 33-37.
12. Witkowski, J. A. (1988) *Trends Genet.* 4, 27-30.
13. Hansen, M. F. & Cavenee, W. K. (1987) *Cancer Res.* 47, 5518-5527.
14. Solomon, E., Voss, R., Hall, V., Bodmer, W. F., Jass, J. R., Jeffreys, A. J., Lucibello, F. C., Patel, I. & Rider, S. H. (1987) *Nature (London)* 328, 616-619.
15. Law, D. J., Olschwang, S., Monpezat, J.-P., Lefrançois, D., Jagelman, D., Petrelli, N. J., Thomas, G. & Feinberg, A. P. (1988) *Science* 241, 961-965.
16. Larsson, C., Skogseid, B., Öberg, K., Nakamura, Y. & Nordenskjöld, M. (1988) *Nature (London)* 332, 85-87.
17. Mathew, C. G. P., Smith, B. A., Thorpe, K., Wong, Z., Royle, N. J., Jeffreys, A. J. & Ponder, B. A. J. (1987) *Nature (London)* 328, 524-526.
18. Noll, W. W. & Collins, M. (1987) *Proc. Natl. Acad. Sci. USA* 84, 3339-3343.
19. Myers, R. M., Lumelsky, N., Lerman, L. S. & Maniatis, T. (1985) *Nature (London)* 313, 495-498.
20. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
21. Kanazawa, H., Noumi, T. & Futai, M. (1986) *Methods Enzymol.* 126, 595-603.
22. Shimosato, Y., Kameya, T. & Hirohashi, S. (1979) *Pathol. Annu.* 14, 215-257.
23. Blin, N. & Stafford, D. M. (1976) *Nucleic Acids Res.* 3, 2303-2308.
24. Sekiya, T., Tokunaga, A. & Fushimi, M. (1985) *Jpn. J. Cancer Res.* 76, 787-791.
25. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
26. Shiraishi, M., Morinaga, S., Noguchi, M., Shimosato, Y. & Sekiya, T. (1987) *Jpn. J. Cancer Res.* 78, 1302-1308.
27. Shiraishi, M. & Sekiya, T. (1988) *Proc. Jpn. Acad.* 64, 25-28.
28. Cavenee, W., Leach, R., Mohandas, T., Pearson, P. & White, R. (1984) *Am. J. Hum. Genet.* 36, 10-24.
29. Ropers, H. H., Gedde-Dahl, T., Jr., & Cox, D. W. (1987) *Cytogenet. Cell Genet.* 46, 213-241.
30. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
31. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* 12, 7035-7056.
32. Sekiya, T., Fushimi, M., Hori, H., Hirohashi, S., Nishimura, S. & Sugimura, T. (1984) *Proc. Natl. Acad. Sci. USA* 81, 4771-4775.
33. Sekiya, T., Fushimi, M., Hirohashi, S. & Tokunaga, A. (1985) *Jpn. J. Cancer Res.* 76, 555-558.
34. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. (1983) *Nature (London)* 302, 33-37.
35. Feinberg, A. P., Vogelstein, B., Droller, M. J., Baylin, S. B. & Nelkin, B. D. (1983) *Science* 220, 1175-1177.
36. Li, H., Gyllenstein, U. B., Cui, X., Saiki, R. K., Erlich, H. A. & Arnheim, N. (1988) *Nature (London)* 335, 414-417.

Single base pair mutation analysis by PNA directed PCR clamping

Henrik Ørum*, Peter E. Nielsen¹*, Michael Egholm², Rolf H. Berg³, Ole Buchardt³ and Christopher Stanley*

PNA Diagnostics A/S, Lersø Park Allé 42, DK-2100 Copenhagen Ø, ¹Research Center for Medical Biotechnology, Department of Biochemistry B, The Panum Institute, Blegdamsvej 3C, DK-2200 Copenhagen N, ²Department of Organic Chemistry, The H.C. Ørsted Institute, Universitetsparken 5, DK 2100 Copenhagen Ø and ³Polymer Group, Materials Department, Risø National Laboratory, DK-4000 Roskilde, Denmark

Received September 16, 1993; Accepted October 19, 1993

ABSTRACT

A novel method that allows direct analysis of single base mutation by the polymerase chain reaction (PCR) is described. The method utilizes the finding that PNAs (peptide nucleic acids) recognize and bind to their complementary nucleic acid sequences with higher thermal stability and specificity than the corresponding deoxyribooligonucleotides and that they cannot function as primers for DNA polymerases. We show that a PNA/DNA complex can effectively block the formation of a PCR product when the PNA is targeted against one of the PCR primer sites. Furthermore, we demonstrate that this blockage allows selective amplification/suppression of target sequences that differ by only one base pair. Finally we show that PNAs can be designed in such a way that blockage can be accomplished when the PNA target sequence is located between the PCR primers.

INTRODUCTION

A multitude of human genetic diseases result from single base mutations in specific genes (1). To facilitate the *in vitro* analysis of such mutations several techniques have been devised. These include enzymatic (2) or chemical (3-4) probing of mismatch complexes, gradient gel electrophoresis (5), use of nucleotide analogues (6) hybridization with allele specific oligonucleotide probes (7) and the oligonucleotide ligation assay (8). To enhance the sensitivity of these methods the target nucleic acid is normally amplified to detectable quantities by the polymerase chain reaction (PCR) (9).

PCR itself has also been used to analyse directly single base mutations by using allele specific oligonucleotides as amplification primers (10-12). Unfortunately, the general applicability of this approach is limited by the fact that the majority of primer-template mismatches have no significant effect on the amplification process (13).

We recently found that PNA (Peptide Nucleic Acid) is a potent DNA mimic in terms of sequence specific hybridization, and obtained results showing that at physiological ionic strength PNA/DNA duplexes are generally 1°C per base pair more stable thermally than the corresponding DNA/DNA duplexes (14-17). Furthermore, our results indicated that the base pair mismatch discrimination is greater for PNA/DNA than for the corresponding DNA/DNA duplexes. In the special case of homopyrimidine PNA, (PNA)₂/DNA triplexes are formed of unprecedented thermal stability and sequence discrimination with complementary oligonucleotides. For example, the complex of PNA T₁₀ with dA₁₀ exhibits a T_m of 76°C, with ΔT_m's for base mismatches ranging between 10-13°C (18). Similarly, T_m for a PNA (T₄CT₅)₂/dA₄GA₅ complex is 79°C, with ΔT_m's for base mismatches ranging between 30-35°C (18).

Taking advantage of these unique properties of PNA, and the fact that PNA cannot function as a primer for DNA polymerase, we now report that PNA can be used to block a PCR amplification process in a sequence specific manner. Furthermore, we show that the specificity of this approach, termed 'PCR clamping', is such that two alleles which differ by only one base pair can be discriminated. Thus this technique allows for direct analysis of single base mutations by PCR.

MATERIALS AND METHODS

The PNAs H-T₁₀-LysNH₂, H-T₅CT₄-LysNH₂, PNA62 (H-TG-TACGTCACTAACTA-NH₂) and PNA176 H-GATCCTGTAC-GTCACTAACTA-NH₂ were synthesized as described (15-17). The plasmid pT10KS was constructed by cloning the complementary oligonucleotides 5'-GATCCT₁₀G and 5'-GATCCA₁₀G into the *Bam*HI site of the Bluescript KS⁺ plasmid (Stratagene). The plasmid pT9C was constructed by cloning the complementary oligonucleotides 5'-TCGACT₅CT₄G and 5'-TCGACA₄GA₅G into the *Sa*I site of pUC19. The plasmid p62-1 was constructed by first cloning the complementary oligonucleotides 5'-GATC-

* To whom correspondence should be addressed

CTGTACGTCACAACCTA-3' and 5'-GATCTAGTTGTGACGTACAG-3' into the *Bam*HI site of pUC19 to obtain p62, followed by cloning of a 556bp *Pst*I/*Hind*III fragment from the phage λ genome into the *Pst*I/*Hind*III site of p62. The plasmids p62-A-KS, p62-T-KS and p62-C-KS were isolated from a mini-library constructed by cloning the degenerate, complementary oligonucleotides 5'-TCGACTCTAGAGGATCTAGTTGTGANGTACAG-3' and 5'-GATCCTGTACNTCACAACCTAGATCCTCTAGAG-3' into the *Sal*I/*Bam*HI site of bluescript KS⁺ (Stratagene). The control plasmids pCKS and PCKS-1 were Bluescript KS⁺ derivatives which do not contain a target sequence for any of the PNAs used in this study. Using standard techniques (19) plasmids were isolated from selected clones of recombinant *E. coli* JM103, purified by buoyant density centrifugation in CsCl gradients and sequenced by the dideoxy method.

The following oligonucleotide primers were used in the PCR reactions: reverse primer (5'-GAAACAGCTATGAC-3'), reverse-1 primer (5'-CACACAGGAAACAGCTATGAC), forward primer (5'-GTAAAACGACGGC-3'), forward-1 primer (5'-GTAAAACGACGGCCAGT), proximal primer (5'-TACCGGGGATC-3') and primers specific for each of the p62 plasmids: p62-1 primer (5'-TGTACGTCACAACCTA-3'), p62-A-1 primer (5'-TGTACATCACAACCTA-3'), p62-A-2 primer (5'-CCTGTACATCACAACCTA-3'), p62-A-3 primer (5'-ATCCTGTACATCACAACCTA-3'), p62-A-4 primer (5'-GATCCTGTACATCACAACCTA-3'), p62-A-5 primer (5'-GTGGATCCTGTACATCACAACCTA-3'), p62-T primer (5'-GGA-TCCTGTACTTCACAACCTA-3') and p62-C primer (5'-GGA-TCCTGTACCTCACAACCTA-3').

PCR amplifications were carried out in a 50 μ l volume containing 0.1 μ g of each plasmid, 0.2 μ M of each primer, 200 μ M dNTP and buffer (10mM Tris-HCl, pH 8.3 (at 25°C), 10mM KCl, and 3mM MgCl₂). The PCR reactions were overlaid with 2 drops of paraffin oil and incubated at 96°C for 2 minutes before the amplification process was initiated by the addition of 3U of the Stoffel polymerase (Perkin Elmer Cetus) or 1U of the supertaq polymerase (AH Diagnostics). When using the supertaq polymerase the buffer was changed to (50mM Tris-HCl, pH 9.0 (25°C), 50mM KCl, 7mM MgCl₂, 16mM (NH₄)₂SO₄ and 0.2mg/ml BSA). Experiments were carried out using either a Minicycler™ (MJ Research) amplifier machine or a LEP amplifier machine (IgG Biotech). Comparative results were obtained independent of the machine and polymerase used. PCR cycle profiles and concentrations of PNAs were as indicated in the figure legends.

T_m values for PNA/DNA DNA/DNA duplexes were determined spectrophotometrically at 260nm in 10mM Na-phosphate, 150mM NaCl and 1mM MgCl₂.

RESULTS

PNAs can effectively block the formation of a PCR product containing a complementary target sequence

Given the higher thermal stability of a PNA/DNA duplex compared to the corresponding DNA/DNA duplex we speculated that PNA might be able to block PCR in a sequence specific manner if targeted against one of the PCR primer sites. Clearly, for such a blocking mechanism to work the PNA must compete effectively against its cognate PCR primer in binding to their common recognition site. To facilitate this requirement, the normal 3 step PCR cycle was expanded with a distinct PNA

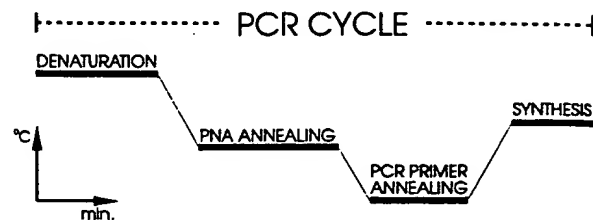


Figure 1. Schematic representation of the PCR cycle profile used in PNA directed clamping.

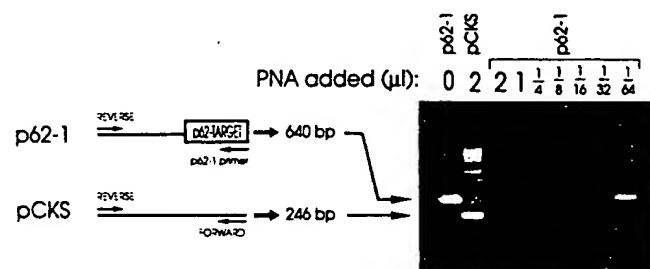


Figure 2. Experimental setup and result of a PCR clamping experiment in the presence of increasing concentrations of PNA62. Lane 1: amplification of the p62-1 plasmid in the absence of PNA62. Lane 2: amplification of the pCKS control plasmid (plasmid containing no PNA62 target) in the presence of 17.8 μ M PNA62. Lanes 3-9: amplification of the p62-1 plasmid in the presence of 17.8 μ M (3), 8.9 μ M (4), 2.2 μ M (5), 1.1 μ M (6), 0.6 μ M (7), 0.3 μ M (8) and 0.15 μ M (9) PNA62. PCR cycle conditions were 96°C, 2min-65°C, 1min-40°C, 30sec-60°C. 2min-30 cycles.

annealing step which 1) precedes the PCR primer annealing step and 2) is set at a temperature that allows only the PNA to bind to its target sequence (Figure 1).

Figure 2 shows the experimental setup and result of a PCR clamping experiment in the presence of increasing amounts of a 15mer PNA, PNA62 (H-TGTACGTCACAACCTA-NH₂). Two plasmid templates were used: the p62-1 plasmid which directs the amplification of a 640bp fragment containing a PNA62 target site and the control plasmid, pCKS, which directs the amplification of a 246bp non-target fragment. When PNA62 is either absent (lane 1) or present at a concentration of 0.15 μ M (lane 9) the p62-1 plasmid directs the synthesis of the expected 640bp PCR fragment. At concentrations at or above 0.3 μ M PNA62, however, no product is produced (lanes 3 to 8). The absence of product is not due to a non-specific inhibitory effect of PNA62 on PCR, since even at the highest concentration used (17.8 μ M) PNA62 will not inhibit the amplification of the expected 246bp fragment from the pCKS control plasmid (lane 2). Furthermore, the ability to clamp PCR is not the result of some unique property of PNA62 since similar results could be obtained with other mixed sequence PNAs (data not shown).

Clamping can be accomplished when the PNA target site is located between the two PCR primers

We next analysed whether the PNA would be able to clamp PCR independent of the relative position of the PNA and PCR primer target sites. We compared the overlap of PNA and PCR primer target sites to the situations where the PNA target site is either 1) located adjacent to a PCR primer site or 2) located in the

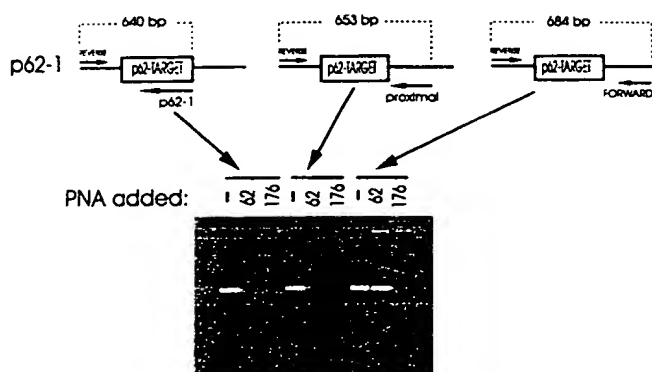


Figure 3. Experimental setup and effect on PNA62 directed clamping of changing the relative position of the PNA and PCR primer target sites. The 3'-end of the forward primer is located 31bp downstream of the PNA62 target site and 26bp downstream of the PNA176 target site. The 3'-end of the proximal primer is located 1bp downstream of the PNA62 target and overlaps the PNA176 by 4bp. The p62-1 primer exactly overlaps the PNA62 target site. Lanes 1-3: amplification of the p62-1 plasmid with reverse and p62-1 primers in the absence (1) or presence of 17.8 μ M of PNA62 (2) or 17.8 μ M of PNA176 (3). Lanes 4-6: amplification of the p62-1 plasmid with reverse and proximal primers in the absence (4) or presence of 17.8 μ M of PNA62 (5) or 17.8 μ M PNA176 (6). Lanes 7-9: amplification of the p62-1 plasmid with reverse and forward primers in the absence (7) or presence of either 17.8 μ M PNA62 (8) or 17.8 μ M PNA176 (9). PCR cycle conditions were 96°C, 2min-65°C, 1min-40°C, 30sec-60°C, 2min-30 cycles.

middle of the PCR region. There are fundamental differences in the underlying mechanism of clamping in these three cases. When the PNA and PCR primer target sites overlap, clamping operates by 'primer exclusion'. Conversely, when the target site is located at a distance from the PCR primer sites, clamping is expected to operate by preventing read-through by the Taq polymerase ('elongation arrest'). Finally, when the PNA target is located adjacent to the PCR primer site, clamping is likely to operate by either preventing polymerase access to the PCR primer and/or by preventing initiation of primer elongation.

Figure 3 shows the experimental setup and the result of changing the relative position of the PNA and PCR primer target sites. Using PNA62, clamping can be accomplished efficiently when the PNA target site either overlaps (lane 2) or is located adjacent to a PCR primer site (lane 4). However, when the PNA target site is located at a distance from the PCR primer site no clamping is observed after 30 cycles (lane 8) suggesting that this PNA/DNA complex is unable to prevent read-through by the polymerase. To test whether an extended PNA62, with an increased T_m for its complementary DNA target, was capable of clamping we synthesized PNA176 (H-GATCCTGTACGTC-ACAACTA-NH₂) which is complementary to the PNA62 target plus the first 5 flanking base pairs in the plasmid. As shown in Figure 3, PNA176 efficiently clamps the PCR process independent of the position of the PNA target site; lane 3: overlapping PNA and PCR primer target sites, lane 6: proximal PNA and PCR primer target sites and lane 9: widely spaced PNA and PCR primer target sites. In experiments with another DNA target sequence we have found that a shorter PNA than PNA62, with a correspondingly lower T_m , is successful in clamping its cognate PCR when its target site is located at a distance from the PCR primers (data not shown). Thus, it would appear that this clamping ability may be a complex function of affinity and kinetics of dissociation.

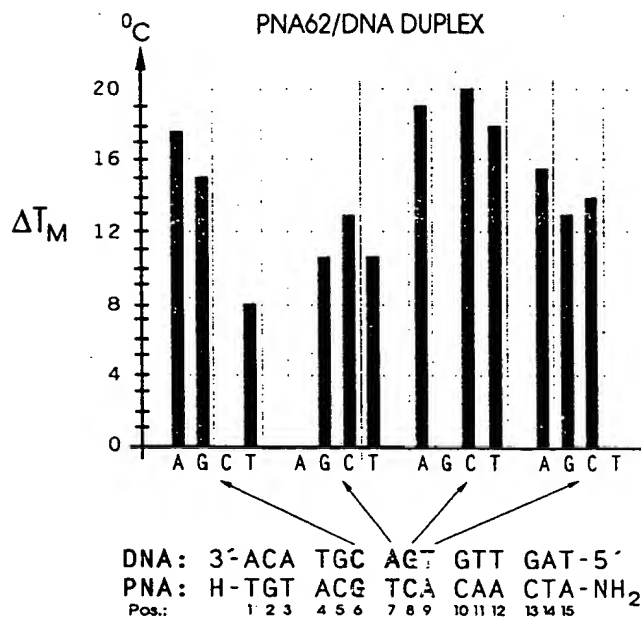


Figure 4. Schematic representation of the effect on T_m of introducing single base mismatches in a PNA62/DNA complex. A series of anti-parallel DNA oligonucleotides were synthesized which contained different single base mismatches to the PNA62 at position 6-9. The T_m value of the fully complementary PNA/DNA duplex is 69°C. The ΔT_m values shown in the Figure indicate the reduction in thermostability that results from the introduction of single base mismatches in the helix.

PNA directed clamping can be used to analyse single base mutations

To explore the potential of the method we tested whether PNA62 would be able to discriminate between fully complementary and single base mismatch targets in a mixed target PCR (i.e. where both targets are present in the PCR reaction mix). As shown in Figure 4 single base mismatches in the PNA62/DNA duplex lower the thermostability of the complex by 8-20°C depending on the type of mismatch and its position in the duplex. Based on these data we chose to analyse mutations at position 6 since these mutations span the largest temperature interval and also include the mutation that exhibits the least helix destabilizing effect, i.e. the PNA G/T DNA mutation (ΔT_m of 8°C).

The configuration chosen for the point mutation analysis was primer exclusion. The PCR reaction mix contained the p62-1 wildtype plasmid, the appropriate mutant p62-plasmid, primers specific for the wildtype and the mutant plasmids and the common reverse primer. Figure 5 shows the experimental setup and the result of the PNA G/T DNA mutation analysis. In the absence of PNA62 two PCR products of sizes corresponding to amplification of the p62-1 and p62-A-KS are produced (lanes 1, 3, 5, 7 and 9). In the presence of PNA62, however, the synthesis of a PCR product is dependent on the size of the mutant primer. When the mutant primer has a size similar to PNA62, addition of the PNA oligomer will suppress the amplification from both wildtype and mutant plasmid (lane 2). This is because the duplex between the mutant primer and its complementary target is less thermostable ($T_m=52^\circ\text{C}$) than the mismatched PNA62/mutant duplex ($T_m=61^\circ\text{C}$, Fig. 4). However, as the size of the mutant primer is increased its ability to compete with PNA62 for binding to its target sequence increases. Thus, at a

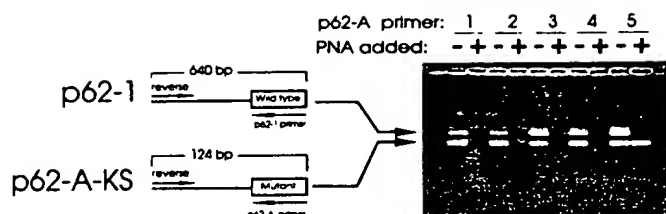


Figure 5. Optimization of the size of the mutant primer required to carry out selective PNA62 directed clamping of the wildtype p62-1 plasmid in the presence of the p62-A-KS single base mutated plasmid. Each reaction contains the p62-1 and p62-A-KS plasmid, the p62-1 primer, one of the p62-A-1 to 5 primers, the common reverse primer and 8.9 μ M PNA62. Lanes 1–2: amplification using the p62-A-1 primer in the absence (1) or presence (2) of PNA62. Lanes 3–4: amplification using the p62-A-2 primer in the absence (3) or presence (4) of PNA62. Lanes 5–6: amplification using the p62-A-3 primer in the absence (5) or presence (6) of PNA62. Lanes 7–8: amplification using the p62-A-4 primer in the absence (7) or presence (8) of PNA62. Lanes 9–10: amplification using the p62-A-5 primer in the absence (9) or presence (10) of PNA62. PCR cycle conditions were 96°C, 2min–65°C, 1min–40°C, 30sec–60°C, 2min–30 cycles.

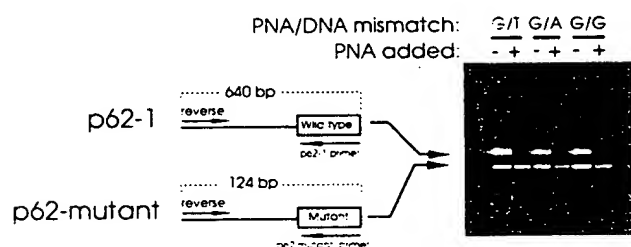


Figure 6. Point mutation analysis with PNA62. Each PCR reaction contains the p62-1 plasmid and one of the single base mutated plasmids p62-A-KS, p62-T-KS or p62-C-KS, the common reverse primer, the appropriate allele specific primers (p62-1 primer, p62-A-5 primer, p62-T-4 primer, p62-C-4 primer) and 8.9 μ M PNA62. Lanes 1–2: co-amplifications of the p62-1 and p62-A-KS plasmids in the absence (1) and presence (2) of PNA62. Lanes 3–4: co-amplifications of the p62-1 and p62-T-KS plasmids in the absence (3) and presence (4) of PNA62. Lanes 5–6: co-amplifications of the p62-1 and p62-C-KS plasmids in the absence (5) and presence (6) of PNA62. PCR cycle conditions were 96°C, 2min–65°C, 1min–40°C, 30sec–60°C, 2min–30 cycles.

mutant primer size of +8 nucleotides (relative to the 15mer PNA62) the p62-A-KS plasmid directs the amplification of a small amount of PCR product (lane 8) and at a mutant primer size of +10 this amplification product is readily visible in the gel (lane 10). Even at a size of +10, however, the mutant primer will not prevent PNA62 from clamping its wildtype target as shown by the lack of the 640bp band in lane 10.

Using a similar experimental setup we then determined the optimal size for the T-mutant and C-mutant primers. These data are compiled in Figure 6 which shows that PNA62 is able to carry out selective suppression of its fully complementary sequence in the presence of all 3 possible point mutations at the position analysed; lane 2: PNA G/T DNA mismatch, lane 4: PNA G/A DNA mismatch and lane 6: PNA G/G DNA mismatch.

Clamping with homopyrimidine PNAs.

In the examples described above the PNAs contained both purine and pyrimidine nucleobases. Such mixed sequence PNAs form highly thermostable duplexes in a preferred anti-parallel

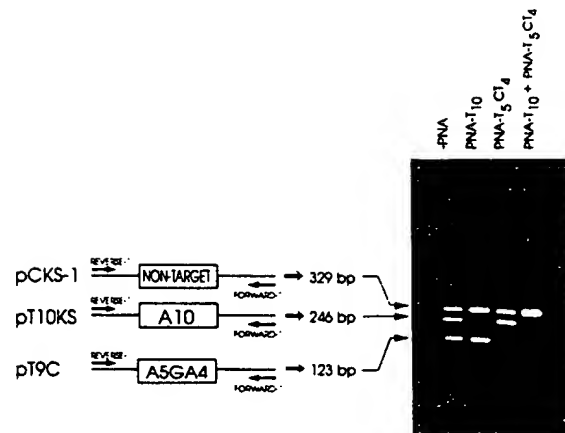


Figure 7. Single base mismatch analysis with two homopyrimidine PNAs complementary to a target sequence located at a distance from the PCR primers. Each PCR reaction contains the pT10KS, pT9C and pCKS-1 (control) plasmids and PNAs as indicated. Lane 1: co-amplifications in the absence of PNAs. Lane 2: co-amplifications in the presence of 3.3 μ M PNA H-T₁₀-LysNH₂. Lane 3: co-amplifications in the presence of 13.2 μ M PNA H-T₄CT₄-LysNH₂. Lane 4: co-amplifications in the presence of 3.3 μ M H-PNA-T₁₀-LysNH₂ and 13.2 μ M PNA H-T₄CT₄-LysNH₂. PCR cycle conditions were 96°C, 2min–62°C, 3min–40°C, 1min–65°C, 2min–35 cycles.

orientation with their target DNA sequence (17). In contrast to this binding mode, homopyrimidine PNAs form extremely thermostable (PNA)₂/DNA triplexes with a preference for parallel orientation with their target DNA sequence (15–16). We therefore also wished to study the PCR clamp technique with such triplex forming PNAs. To obtain thermal stabilities comparable to the previously described PNA62/DNA duplex we chose two 10mer PNAs (PNA H-T₁₀-LysNH₂, T_m =75°C and H-T₄CT₄-LysNH₂, T_m =79°C) which differ from each other at a single base position. Both of these PNAs acted as efficient and sequence specific clamps in a PCR amplification process and both PNAs were very efficient in blocking their cognate PCR process independent of the location of the PNA and PCR primer sites (data not shown). Furthermore, as shown in Figure 7 both homopyrimidine PNAs were able to discriminate between their fully complementary and single base mismatch targets when the PNA target site is located at a distance from the PCR primer sites. In the absence of either of the PNAs three PCR products of sizes corresponding to amplification of the pT10KS, pT9C and pCKS-1 control plasmids are visible in the gel (Figure 4, lane 1). If PNA H-T₁₀-LysNH₂ is included in the PCR reaction alone the products corresponding to amplification of the pT9C and pCKS-1 control plasmids are seen (lane 2). Similarly, PNA H-T₄CT₄-LysNH₂ suppresses the amplification of its cognate target fragment, whilst leaving amplification of the pT10KS and pCKS-1 control plasmids unaffected (lane 3). In the presence of both PNAs only the PCR product corresponding to the pCKS-1 control plasmid is seen.

DISCUSSION

We have used PNA to develop a method that converts a PCR amplification process into an efficient analytical tool for the direct detection of single base mutations. In our hands the method is very robust. The two different modes of PCR blocking by PNA

(primer exclusion and elongation arrest) and the two different ways in which PNA target recognition can occur (duplex vs. triplex), further provide great versatility and flexibility to the PNA/PCR clamp system.

It is interesting to note that clamping can operate efficiently even with incomplete binding of PNA to its DNA targets. For example, our calculations show that if 1% of all target sequences escape clamping in *each* cycle the maximum amplification factor after 30 cycles is only 9-fold, which will not generally produce a detectable signal on a gel. Indeed, in PCR amplifications of genomic material we predict that as much as 10% of the target sequences can escape clamping without generating a detectable signal (equivalent to a maximum amplification factor of 2500 in 30 cycles).

In a PNA clamping protocol with mixed sequence PNAs we prefer to use the primer exclusion principle for the following reasons. First, this clamping mode places the least physical demands on the PNA, i.e. clamping does not require that the PNA, once bound to its target, is able to prevent read-through by the polymerase as is the case in the elongation arrest clamping mode. Second, the only variables in the primer exclusion clamping mode are the T_m of the PNA and the PCR primers and these can be tuned to precision simply by changing either the sizes of the PNA and PCR primers, or by altering their exact position on the target DNA. Third, when using the primer exclusion principle there is the further advantage that, in addition to blocking its cognate target site, the PNA will compete with the PCR primer for any cryptic primer sites in the genome, thereby suppressing any occurrence of non-specific background in the PCR process directed by this primer.

In order to target unique sequences in the human genome a primer of at least 17bp is usually required, the T_m 's of which typically range between 50–60°C. Thus, for the successful projection of our PNA clamping approach to the analysis of point mutations in the human genome, PNAs with T_m 's above 60°C must be able to effectively discriminate between their fully complementary and single base mismatched target DNA. Using a mixed sequence 15mer PNA with a T_m of 69°C we have shown, in a model system, that three different point mutations at a single position can be discriminated, suggesting that PNA clamping can be used as an effective diagnostic tool for the analysis of mutations in complex genomes. We acknowledge that only three out of twelve possible mismatches have been analysed and only in a single sequence context. However, we have shown in the present study that the PCR clamp system can efficiently discriminate the most difficult case in our system (PNA G/T DNA mismatch with a ΔT_m of only 8°C). We are therefore confident that, given the great flexibility of the method, conditions can be found to discriminate any single point mutation. When such mutations are present at low frequency it is interesting to speculate that the PNA clamp may instead be directed against the large excess of non-mutated genes the presence of which leads to unwanted background in diagnostic procedures. We now intend to apply the PCR clamp technique to the analysis of point mutations in human, animal and plant genetic material.

REFERENCES

- Landegren, U., Kaiser, R., Caskey, C.T. and Hood, L. (1988) *Science* **242**, 229–237.
- Myers, R.M., Larin, Z. and Maniatis, T. (1985) *Science* **230**, 1242–1246.
- Novack, D.F., Casna, N.J., Fisher, S.G. and Ford, J.P. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 586–590.
- Cotton, R.G.H., Rodrigues, N.R. and Campbell, D.R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4397–4401.
- Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. (1985) *Nature* **313**, 495–498.
- Kornher, S. and Livak K.J. (1989) *Nucleic Acid Res.* **17**, 7779–7784.
- Saiki, R.K., Walsh, P.S., Levenson, C.H. and Erlich, H.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6230–6234.
- Landegren, U., Kaiser, R., Saunders, J. and Hood, L. (1988) *Science* **241**, 1077–1080.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science* **239**, 487–491.
- Ehlen, T. and Dubeau, L. (1989) *Biochem. Biophys. Res. Commun.* **160**, 441–447.
- Gibbs, R.A., Nguyen P. and Caskey, C.T. (1989) *Nucleic Acid Res.* **17**, 2437–2448.
- Chehab, F.F. and Kan Y.W. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9178–9182.
- Kwok, G., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J. (1990) *Nucleic Acid Res.* **18**, 999–1005.
- Nielsen, P.E., Egholm, M., Berg, R.H. and Buchardt, O. (1991) *Science* **254**, 1497–1500.
- Egholm, M., Buchardt, O., Nielsen, P.E. and Berg, R.H. (1992) *J. Amer. Chem. Soc.* **114**, 1895–1897.
- Egholm, M., Buchardt, O., Nielsen, P.E. and Berg, R.H. (1992) *J. Amer. Chem. Soc.* **114**, 9677–9678.
- Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B. and Nielsen, P.E. (1992) *Nature* (in press).
- Egholm, M., Behrens, C., Christensen, L., Berg, R.H., Nielsen, P.E. and Buchardt, O. (1993) *J. Chem. Soc. Chem. Commun.* pp.800–801.
- Sambrook, J., Fritsch, E. and Maniatis, T. (1982). *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

A High-Resolution Microsatellite Map of the Mouse Genome

Michael Rhodes,¹ Richard Straw,¹ Supem Fernando,¹ Andrew Evans,¹
Tregaye Lacey,¹ Andrew Dearlove,¹ John Greyststrong,¹ Joanne Walker,¹
Paula Watson,¹ Paul Weston,¹ Maria Kelly,¹ Dilip Taylor,¹ Keith Gibson,¹
Chris Mundy,¹ Franck Bourgade,² Christophe Poirier,²
Dominique Simon,² Ana Lucia Bueno Brunialti,² Xavier Montagutelli,²
Jean-Louis Guénet,² Andy Haynes,³ and Steve D.M. Brown^{3,4,5}

¹United Kingdom Human Genome Mapping Project (HGMP) Resource Centre, Hinxton CB10 1RQ, UK; ²Institut Pasteur, Paris, France; ³Medical Research Council Mouse Genome Centre and ⁴Medical Research Council Mammalian Genetics Unit, Harwell OX11 ORD, UK

The European Collaborative Interspecific Backcross (EUCIB) resource was constructed for the purposes of high-resolution genetic mapping of the mouse genome (Breen et al. 1994). The large *Mus spretus*/C57BL/6 backcross of 982 progeny has a genetic resolution of 0.3 cM at the 95% confidence level (~500 kb in the mouse genome). We have used the EUCIB mapping resource to develop a genome-wide high-resolution genetic map incorporating 3368 microsatellites. The microsatellites are distributed among 2302 genetically separated bins with 1.46 markers per bin on average. Average bin separation is 0.61 cM. This high-resolution genetic map will aid the construction of a robust physical map of the mouse genome.

The mouse is a pivotal model organism for the genome program and with its battery of mutagenic, transgenic, and developmental biology approaches is set to play a key role as mammalian genetics moves from genomics to studies of gene function (Copeland et al. 1993; Dietrich et al. 1995; Brown and Peters 1996). The development of genetic and physical maps in the mouse is an important step toward providing the genome resources for future gene function studies (Dietrich et al. 1995). The construction of a high density map of mouse simple-sequence polymorphisms at intermediate resolution is complete (Dietrich et al. 1996). The future development of genome-wide physical maps (Hudson et al. 1996) will assist gene mapping as well as providing the clone resources for gene identification. In addition, the map will provide the substrates for preparing sequence-ready maps for comparative sequencing, which will itself speed the process of gene identification. Physical maps will also underpin the development of comprehensive, high-resolution gene maps (Schuler et al. 1996) that can be used for the characterization of mouse mutations

and provide a new cache of gene sequences that can be related to loci on the human genome by the conserved linkage groups identified between the two species (Copeland et al. 1993; Andersson et al. 1996). Nevertheless, there are no complete physical maps yet available for any mouse chromosome. The development of a high-resolution genetic map can enhance the production of a robust physical map on any mouse chromosome.

The ability to undertake large genetic crosses between defined mouse strains means the construction of high-resolution genetic maps can be readily achieved. Most notably, large interspecific or intersubspecific backcrosses between laboratory strains of mice and wild species such as *Mus spretus* or *Mus castaneus* (Avner et al. 1988) has transformed mouse genetic mapping (for review, see Copeland et al. 1993). Large numbers of backcross progeny can be readily derived from such crosses providing the requisite numbers of meioses to achieve high genetic resolution. Additionally, the use of crosses between relatively diverged species contributes to the large numbers of markers that are variant between the parental strains. Approximately 90% of microsatellites show size variation between laboratory strains and the wild species, *M. spretus* and *M. castaneus*

Corresponding author.

E-MAIL s.brown@har.mrc.ac.uk; FAX 01235 824542.

(Dietrich et al. 1992). Small intersubspecific crosses have been used for the construction of genome-wide gene or microsatellite maps at intermediate resolution (Dietrich et al. 1996; see above). Large interspecific backcrosses of a 1000 progeny or more carrying a specific mutation of interest have been used widely for high-resolution genetic mapping of the mutant locus as a route to positional cloning of the gene (Brown 1994, 1996). However, to date there has been no systematic attempt to use the high resolution afforded by large interspecific backcrosses to construct genome-wide high-resolution maps.

Recently, we reported the construction of a high-resolution mouse mapping resource consisting of an interspecific backcross of nearly 1000 progeny (Breen et al. 1994). A backcross of this size has a genetic resolution of 0.3 cM at the 95% confidence level that equates to ~600 kb in the mouse genome. We have now used this backcross to construct a high-resolution and high-density microsatellite map of the mouse genome. This high-resolution genetic map will be the anchor for the construction of a high integrity physical map of the mouse genome.

RESULTS

Identification of Panels of Recombinants from EUCIB for High-Resolution Mapping

We have described the construction of a large interspecific backcross between C57BL/6 and *Mus spretus*—the European Collaborative Interspecific Backcross (EUCIB)—comprising 982 backcross progeny. Backcross progeny were initially typed for 78 primary anchor loci spanning the entire genome with 3–6 anchors per chromosome (Breen et al. 1994). Subsequently, a number of additional anchor markers were added. Where it became apparent from further mapping studies that the proximal and distal anchors available on a particular chromosome did not represent either the most centromeric or telomeric markers, additional anchors were added for the mapping of markers close to the centromere and telomere. The anchor map identifies the great majority of backcross progeny mice recombinant in any interanchor interval (excluding only those rare individuals that have double interanchor recombinants) and provides panels of mice for high-resolution mapping in each interanchor chromosome region. In total, 93 primary anchors were assigned (see Table 1). Subsequently, a large number of secondary anchors (principally microsatellite markers) were added to the map reducing the size of

recombinant panels still further and allowing for rapid high-resolution mapping of markers in any chromosome region. Final interanchor intervals comprised panels of ~36 recombinant mice on average and thus corresponded to a genetic interval of ~4 cM.

Mapping Microsatellites at High-Resolution on the EUCIB Backcross

To develop the high resolution genetic map, a large number of microsatellites markers from the Whitehead/MIT map (Dietrich et al. 1996) were analyzed through the EUCIB backcross. The bulk of microsatellite mapping used a novel, high-throughput and semiautomated fluorescent dUTP genotyping approach (Rhodes et al. 1997)—2278 of the total of 3368 were added to the map by this approach. The remainder were mapped either by use of standard agarose gel electrophoresis or alternatively with an enhanced chemiluminescence approach (Vignal et al. 1993). Following the determination of the parental allele sizes (C57BL/6 and *M. spretus*), the appropriate recombinant panel of mice was genotyped. Given the limited resolution afforded by previous maps, it was not always apparent which interanchor interval a microsatellite would lie within and, therefore, which recombinant panel should be typed. Under these circumstances, appropriate adjacent panels were typed.

Although we tested all the Whitehead/MIT primers available during the period of map construction, inevitably a proportion failed to amplify product or proved problematic for reliable scoring and mapping. For the 4450 microsatellite markers tested by the semiautomated fluorescent dUTP genotyping approach, 51.2% amplified and were mapped successfully. Of the 48.8% that failed to be added to the map, 14.7% failed to give any product whatsoever (on either C57BL/6 or *M. spretus* DNA) and 31.1% produced some product but was not scoreable (e.g., multiple bands or variable product sizes). A small percentage, 3.0%, gave reliable, but identical, products between C57BL/6 and *M. spretus* DNA and were therefore not mappable.

Given the high throughput requirements of the project, we did not return to failed primer sets to optimize PCR conditions, and as a consequence, a proportion of microsatellites were not added to the map on each chromosome. Ultimately, of the Whitehead/MIT microsatellites available to us during map construction, 56% were added to the EUCIB map. For individual chromosomes, the proportion of Whitehead/MIT microsatellites placed on

Table 1. Summary of Markers and Map Statistics by Chromosome

Chromosome	Primary anchors	Microsatellites	Percent MIT	Totals	Bins	Length (cM)
1	4	226	48	230	153	90
2	5	331	70	336	197	95
3	4	176	52	180	128	77
4	5	208	64	213	158	84
5	4	225	60	229	137	92
6	6	195	58	201	139	78
7	4	201	62	205	148	71
8	3	172	52	175	127	62
9	6	143	48	149	107	73
10	6	142	53	148	105	68
11	5	142	45	147	97	72
12	5	149	58	154	116	60
13	4	178	61	182	116	68
14	5	130	54	135	101	53
15	3	144	58	147	91	62
16	3	110	55	113	77	56
17	4	117	52	121	90	59
18	7	115	56	122	67	55
19	4	67	57	71	54	54
X	6	104	52	110	94	69
Totals	93	3275	56	3368	2302	1398

The number of primary anchors and microsatellites mapped to each chromosome is indicated. Percent MIT indicates the proportion of Whitehead/MIT microsatellites available to us at the time of mapping that were added to the EUCIB genetic map. In addition, the numbers of marker bins for each chromosome is given along with the genetic map length (the distance between the most proximal and distal anchors mapped—see Methods).

the EUCIB map varied from 45% to 70% of Whitehead/MIT microsatellite markers (see Table 1).

The MBx Database—Construction of the EUCIB High-Resolution Microsatellite Map

The MBx database that supports the EUCIB program has been described previously (Breen et al. 1994). Genotypes were entered into the MBx database and genetic maps were constructed. Determining locus order rather than genetic distance was the primary consideration for the construction of genetic maps because this provides the most important enhancement to future physical maps that will be underpinned by the high-resolution genetic map. The order of microsatellite markers along each chromosome was determined by a haplotype analysis that minimizes the recombinants in any chromosomal region. The genetic distances displayed in the EUCIB Genetic Map and MultiMaps are calculated so that the marker order derived by haplotype analysis

is maintained (see Methods); these displays (see Fig. 1) are available on the World Wide Web site: ([URL:http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html](http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html)), which also includes genotype data for individual markers on the maps. Direct access to the MBx database to view haplotypes is also available (see Methods). The MBx database provides scrollable tables of haplotypes for each chromosome, identifying and highlighting all recombination events. It is possible to select and display all mice containing recombination events in a particular interval to assess the raw data and to evaluate how robust locus order is.

The EUCIB High-Resolution Genetic Map

In total, 3368 microsatellites and anchors have been mapped and ordered at high resolution on the EUCIB backcross (see Table 1). The total length of the high resolution EUCIB genetic map is 1398 cM as calculated from proximal-distal anchor distances on each chromosome (Table 1). On each chromo-

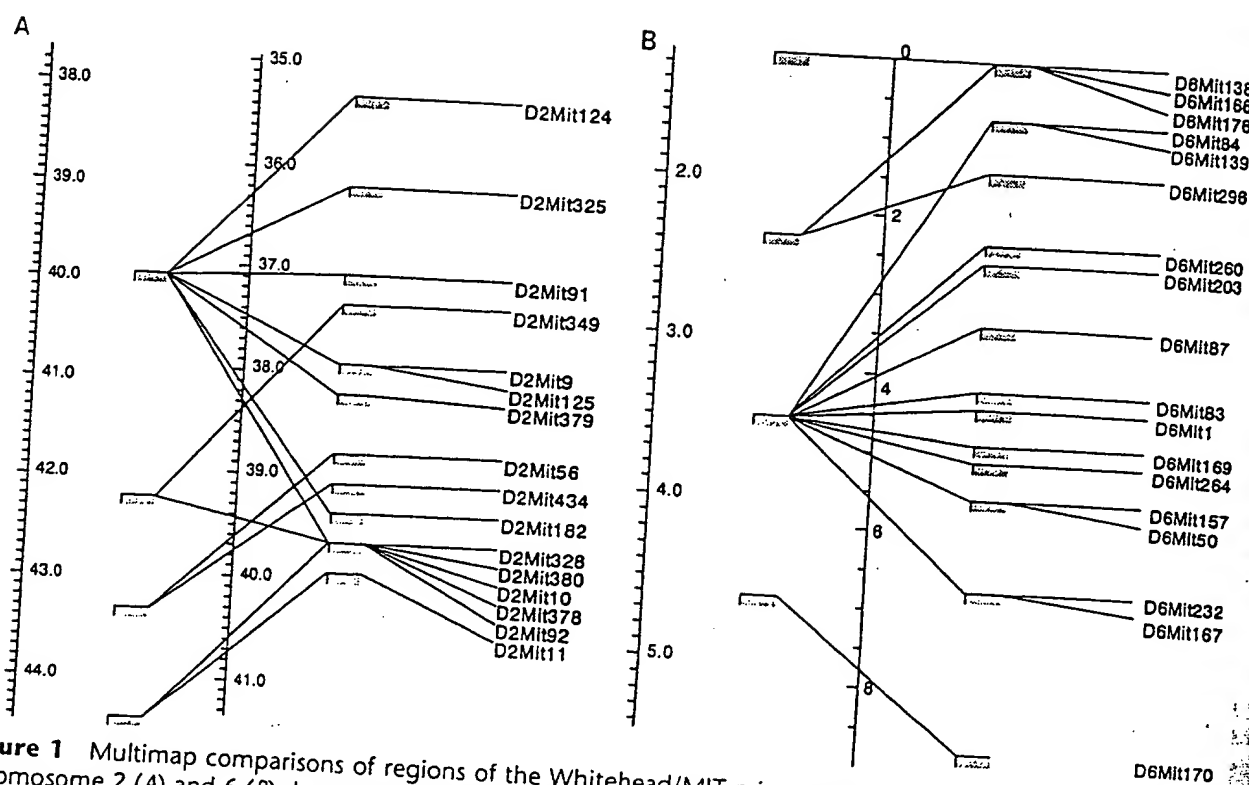


Figure 1 Multimap comparisons of regions of the Whitehead/MIT microsatellite map and the EUCIB map from chromosome 2 (A) and 6 (B) demonstrating increases in bin number and resolution by use of the EUCIB resource. Genetic map positions of Whitehead/MIT bins are shown at left with the genetic map position of the EUCIB bins at right.

ome, a few microsatellites map beyond the most proximal or most distal anchors. However, for these markers it is not feasible to determine accurate anchor-marker distances (see Methods) and for this reason, these markers have not been included in the calculation of total genetic length. Thus, 1398 cM represents the minimum length of the genetic map. The total calculated length of the EUCIB map is very similar to that reported from the genome-wide Freckle C57BL/6 \times *M. spretus* backcross map (Cope and Jenkins 1991) whose genetic length (as reported in Dietrich et al. 1996) is 1385 cM.

The high resolution of the EUCIB map is related by the fact that the microsatellites were distributed among 2302 bins giving 1.46 markers per bin on average—a bin being defined as a minimally resolved genetic interval separated from adjacent bins by proximal and distal recombination breakpoints for all markers in that bin. Table 2 illustrates the distribution of markers per bin. Seventy-one of the bins contain only 1 marker, again highlighting the high resolution of the mapping process. Most notably, the average separation between markers is 0.61 cM. The average intermarker interval is 1.6 Mb or ~800 kb.

Figure 1 shows a multimap comparison of the EUCIB and Whitehead/MIT maps and illustrates the typical increase in resolution afforded by the EUCIB high-resolution mapping resource. It is important to

Table 2. Marker Distribution Across Bins

No. of markers per bin	No. of bins
1	1640
2	427
3	140
4	51
5	29
6	9
7	4
8	0
9	0
10	0
11	1
12	1

The numbers of markers in bins carrying 1, 2, 3, 4 markers, etc., is given.

note when comparing the EUCIB and Whitehead/MIT maps that for the Whitehead/MIT map (Dietrich et al. 1994) constructed from a limited number of F_2 intercross progeny, statistical support for order of a given marker could vary, either because of incomplete genotyping or because the marker is dominant rather than codominant. Thus multimap comparisons of Whitehead/MIT and EUCIB maps illustrate, not unsurprisingly, that markers lying in adjacent bins on the Whitehead/MIT map are sometimes found to interdigitate when their order is determined at high resolution on the EUCIB map (see Fig. 1). For this reason, we have chosen a relatively high cutoff point to assess the frequency of markers that deviate in position between the EUCIB and Whitehead/MIT maps. Proceeding systematically proximal to distal on each chromosome, we co-aligned each EUCIB primary or secondary anchor marker with the corresponding locus on the Whitehead/MIT map, and then assessed each microsatellite lying in the following inter-anchor segment on the EUCIB map for significant deviations with the Whitehead/MIT map. (Only markers lying between the most proximal and distal anchors on each chromosome were subject to this analysis to avoid biases from the few poorly mapped markers lying outside these anchor loci.) Choosing a cutoff point of 10 cM, only 76 markers (2.3%) on the EUCIB map mapped 10 cM or more from their expected location on the basis of the Whitehead/MIT map. At 15 and 20 cM, this figure dropped to 1.1% and 0.8%, respectively. Overall, there is excellent agreement between Whitehead/MIT and EUCIB maps.

Distribution of Markers and Recombination Events

Only two large bins—one of 11 and one of 12 markers (see Table 2)—remain on the final EUCIB map. We examined these bins to see if they corresponded to any of the large bins on the Whitehead/MIT map that may be accounted for by regions of crossover suppression that are common between the two crosses used for mapping (the Whitehead/MIT map was constructed by use of a (C57BL/6J-*ob/ob* \times *Mus castaneus*) F_2 intercross (see Dietrich et al. 1996). The bin of 11 markers is found on chromosome 16 at position 6.34 cM and includes the microsatellites *D16Mit8*, 32, 33, 79, 121, 122, 131, 142, 161, 180, and 181. The bin of 12 markers is found on chromosome 2 at position 32.23 cM and includes the microsatellites *D2Mit8*, 72, 89, 155, 157, 240, 241, 298, 323, 373, 433, and 471. However, neither of these bins corresponded to large, unseparated bins

of markers on the Whitehead/MIT map (Dietrich et al. 1996).

We have also examined the EUCIB maps to determine the largest genetic gaps. Only five gaps >5 cM were identified between the most proximal and distal anchors. On chromosome 1, a gap of 8.59 cM separates *D1Mit65* and *D1Mit118*. On chromosome 5, a gap of 7.38 cM was found to separate *DSMit160* and *DSNds6*, whereas on chromosome 9 two gaps between *D9Mit217* and *D9Mit58* (6.60 cM) and *D9Mit294* and *D9Mit42* (5.44 cM) were identified. On chromosome 18, a gap of 9.68 cM was identified between *D18Mit33* and *D18Mit8*. None of these gaps corresponded to any of the larger genetic intervals on the Whitehead/MIT map (Dietrich et al. 1996; see Discussion).

In total, 17,029 distinct recombination events were observed in the MBx database among the 982 progeny and distributed across all 20 chromosomes (see Table 3). On average, each mouse carries 17.3 recombination events. Forty-four percent of chromosomes did not show an observable recombination event. Forty percent of chromosomes (7863) showed a single recombinant, 7.2% of chromosomes demonstrated double recombinants, and 5.6% triple recombinants. We have used a goodness-of-fit test to analyze the distribution of recombinant classes on each chromosome for fit to Poisson. None of the 20 chromosomes fit a Poisson distribution, differing significantly in all cases (see Table 3). As observed for the original EUCIB anchor map (Breen et al. 1994), and in agreement with other reports (Ceci et al. 1989; Saunders and Seldin 1990; Nadeau et al. 1991; Reeves et al. 1991, 1997), there is a general over-representation of single recombinants, and double recombinants are under-represented on all chromosomes—probably because of crossover suppression. In general, triple recombinant classes and classes carrying larger numbers of recombinants are over-represented—probably largely because of genotyping errors (see below). In total, 10,991 chromosomes (56%) carried one or more crossovers. In general, there was a broad relationship between a chromosome's genetic length and the total number of recombinant events observed for that chromosome. Overall, however, the relationship between numbers of recombinants per chromosome and genetic length across all 20 chromosomes was not significant (see legend to Table 3). A binomial test to identify those chromosomes that contribute significantly fewer or greater recombinants than expected indicates that chromosomes 1, 7, 8, and 18 contribute significantly more recombinants than expected, whereas chromosomes 3, 11,

Table 3. Number and Distribution of Recombinants by Chromosome

Recomb. class	Chromosome										
	1	2	3	4	5	6	7	8	9	10	11
0	316 (282)	332 (281)	447 (418)	350 (318)	312 (291)	376 (342)	443 (371)	395 (347)	433 (414)	448 (400)	438 (490)
1	378 (352)	352 (352)	369 (367)	382 (359)	407 (354)	395 (361)	344 (361)	385 (361)	401 (358)	366 (359)	437 (341)
2	141 (220)	157 (220)	80 (152)	119 (202)	130 (215)	89 (190)	69 (176)	84 (188)	60 (154)	69 (161)	88 (118)
3	81 (91)	67 (92)	57 (43)	71 (76)	72 (87)	71 (67)	76 (57)	66 (65)	50 (44)	56 (48)	17 (27)
4	33 (29)	42 (29)	19 (9)	40 (21)	28 (27)	27 (18)	22 (14)	15 (17)	22 (10)	19 (11)	0 (5)
5	19 (7)	16 (7)	4 (2)	10 (5)	16 (7)	14 (4)	19 (3)	24 (4)	10 (2)	17 (2)	1 (1)
6	8 (2)	11 (2)	3 (1)	7 (1)	10 (1)	5 (1)	5 (1)	6 (1)	3 (1)	2 (2)	0 (1)
7	4 (1)	1 (1)	1 (1)	2 (1)	2 (1)	2 (1)	2 (1)	3 (1)	3 (1)	4 (1)	0 (1)
8	1 (0)	2 (0)	1 (0)	1 (0)	2 (0)	2 (0)	1 (0)	2 (0)	0 (0)	1 (0)	0 (0)
9	0 (0)	0 (0)	1 (0)	0 (0)	1 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)
10	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
11	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
12	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
13	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
14	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
χ^2	101.38	98.9	70.03	89.07	123.72	150.85	282.99	317.6	125.38	128.19	46.81
Total	1225	1229	838	1107	1194	1035	955	1023	848	881	682
recomb. length (cM)	90	95	77	84	92	78	71	62	73	68	72
Binomial test	16.151*	4.78	11.269*	7.302	5.14	8.032	9.9*	99.353*	2.015	3.524	45.725*

Table 3. (Continued)

Recomb. class	Chromosome									Total chr.	Total recombs.	Percent Total Chr.
	12	13	14	15	16	17	18	19	X			
0	462 (475)	475 (450)	578 (547)	503 (552)	507 (557)	474 (530)	467 (457)	483 (505)	410 (425)	8649	0	44.038
1	406 (345)	383 (351)	329 (320)	429 (318)	417 (316)	445 (327)	397 (350)	412 (336)	429 (356)	7863	7863	40.036
2	50 (125)	42 (137)	12 (94)	21 (92)	38 (90)	31 (101)	18 (134)	36 (112)	72 (149)	1406	2812	7.159
3	56 (30)	58 (36)	48 (18)	24 (18)	18 (17)	29 (21)	93 (34)	42 (25)	47 (42)	1099	3297	5.596
4	4 (6)	8 (7)	2 (3)	4 (3)	0 (2)	3 (3)	0 (7)	3 (4)	17 (9)	308	1232	1.568
5	2 (1)	10 (1)	9 (3)	0 (0)	2 (2)	0 (0)	5 (1)	4 (1)	4 (2)	186	930	0.947
6	1	3	3	0	0	0	1	2	1	71	426	0.362
7	1	1	1	1	0	0	0	0	2	30	210	0.153
8	0	1	0	0	0	0	0	0	0	14	112	0.071
9	0	0	0	0	0	0	1	0	0	4	36	0.020
10	0	0	0	0	0	0	0	0	0	5	50	0.025
11	0	1	0	0	0	0	0	0	0	3	33	0.015
12	0	0	0	0	0	0	0	0	0	0	0	0.000
13	0	0	0	0	0	0	0	0	0	1	13	0.076
14	0	0	0	0	0	0	0	0	0	0	0	0.000
15	0	0	0	0	0	0	0	0	0	1	15	0.088
χ^2	78.53	114.52	168.93	101.41	66.55	100.31	208.27	85.23	73.88			
Total												
recomb.	713	767	575	566	557	606	752	654	822		17029	
Length (cM)	60	68	53	62	56	59	55	54	69			
Binomial test	0.456	4.77	8.023	49.609*	23.913*	18.445*	10.46*	0.022	0.428			

The number of nonrecombinant and recombinant classes observed for each chromosome are listed along with the numbers of classes expected according to a Poisson distribution (in parentheses). Where the expected number is <1 , no figure is given. χ^2 values from goodness-of-fit tests applied to each chromosome are given. All chromosomes differ very significantly from a Poisson distribution at $p < 10^{-9}$. Although there appears to be a broad relationship between numbers of total recombinants per chromosome and genetic length (cM), the observed distribution deviates significantly from that expected if the number of recombinants varies in proportion to genetic length ($\chi^2 = 313.64$, $P < 10^{-10}$, 19 df). A binomial test of significance has also been applied to each chromosome to identify those chromosomes that depart most significantly from expected, and the χ^2 values for each chromosome are listed. We have taken account of the likelihood of false positives, i.e., chromosomes that depart from expected by chance. Monte Carlo simulations indicate that when adopting a significance level of 0.05, one false-positive chromosome will arise with a frequency of 0.6317. Reducing the false-positive rate to 0.05 the binomial test needs to be applied at a significance level of 0.002. At $\alpha = 0.002$ chromosomes 1, 3, 7, 8, 11, 15, 16, 17 and 18 show a significant departure from expected (indicated by asterisk).

15, 16, and 17 contribute significantly fewer (see legend to Table 3).

Error Correction

Multiple recombinants observed on an individual chromosome (see above) can be characteristic of genotyping errors. For this reason, we have sought to identify those multiple recombinants that may be most indicative of genotyping errors to reach some assessment of the overall level of genotyping error in the EUCIB dataset. Error correction on each chromosome has been proceeded by the identification of haplotypes 1(.)2(.)1 and 2(.)1(.)2 for adjacent markers within the database where 1 represents a homozygote, 2 a heterozygote and (.) represents a variable number of intervening markers for which genotype information might not be available in any haplotype (see legend to Table 4). These apparently double recombinant haplotypes for very closely linked markers would be expected to occur rarely, if at all, and contribute to a proportion of double and triple recombinant chromosomes and to chromosomes carrying larger numbers of recombinants (see above). Subsequently, having identified these double recombinant haplotypes, primary data entry is checked, or in some cases scorings are repeated. Eventually, when error correction is complete, reordering of markers is carried out by MBx. Nevertheless, following error correction and reordering, some aberrant haplotypes remain (see Table 4 for the total number of spurious double recombinant haplotypes per chromosome). In total, 1158 double recombinants of the form 1(.)2(.)1 or 2(.)1(.)2 remain. Across the whole dataset, 41% of these double recombinants are of the form 121 or 212 with no intervening unscored markers. Sixty-six percent of double recombinants are of the form 121, 212 and 1(.)2(.)1 or 2(.)1(.)2 in which only a single intervening marker (.) is unscored. The high frequency of these apparently closely spaced double recombinants is very suggestive of genotyping errors rather than true recombination events.

In general, if we assume that all aberrant haplotype events represent genotyping errors, then the average genotyping error rate across the whole genome is ~0.01. The rate varies from chromosome to chromosome, therefore, on some chromosomes, for example, chromosome 16, it is as low as 0.002. However, we have undertaken additional analyses to empirically estimate the residual error rate for genotypes within the EUCIB dataset. Additionally, we have attempted to estimate the overall error rate for marker order across the genome.

Table 4. Number and Chromosome Distribution of Aberrant Double and Triple Recombinants

Chromosome	Double recombinants	Triple recombinants
	1(.)2(.)1 2(.)1(.)2	11(.)2(.)1(.)2 22(.)1(.)2(.)1
1	83	8
2	47	4
3	44	18
4	111	16
5	65	1
6	86	19
7	120	13
8	106	9
9	63	10
10	73	7
11	16	2
12	52	6
13	46	4
14	57	5
15	15	10
16	8	0
17	28	2
18	28	2
19	46	7
X	64	17
Totals	1158	160

Double recombinants of the form 1(.)2(.)1 and 2(.)1(.)2 (1 homozygote; 2, heterozygote) that may represent genotyping errors (see text) were identified for each chromosome from the MBx database. (.) represents a variable number of intervening markers with no genotypes for a haplotype. In 10 intervening markers for which there was no genotype information were permitted. Thus, at the limit, double recombinants of genotype 1 (10 markers, no genotypes) 2 (10 markers, no genotypes) 1, and 2 (10 markers, no genotypes) 1 (10 markers, no genotypes) 2 were identified. For chromosomes 1 and 5, a further round of error checking of double recombinant genotypes and reordering took place for the chromosomes to assess the underlying error rate (see text). Triple recombinants of the form 11(.)2(.)1(.)2 and 22(.)1(.)2(.)1, which may represent local misorder (see text), were also identified from MBx. Up to 10 intervening markers (.) for which there is no genotype information were again permitted.

Genotyping Error Rate

Following the final rounds of data production, error checking, and ordering, we chose two chromosomes—1 and 5—and identified all remaining 1(.)2(.)1 and 2(.)1(.)2 haplotypes for adjacent markers from within the dataset. Some 1(.)2(.)1 and 2(.)1(.)2 haplotypes will occur as part of normal

complex triple recombinant haplotypes [e.g., 22(.)1(.)2(.)11] among adjacent markers. These triple recombinant haplotypes can arise because of errors in typing, or more likely, errors in ordering in which the inversion of the central two markers would remove the triple recombinant haplotype and substitute a single recombinant haplotype in its place (see below). Nevertheless, for chromosomes 1 and 5, the aberrant genotypes were rescored in all cases. This involved repeating the appropriate PCR reactions under identical reaction conditions. Following retyping, these chromosomes were reordered. Table 5 gives the reduction in the number of 1(.)2(.)1 and 2(.)1(.)2 haplotypes observed on each chromosome following second rounds of error checking and ordering and, therefore, a more accurate figure of the genotyping error rate. Taking both chromosomes together, of the original aberrant haplotype genotypings, 48% (134) were found to be incorrect, giving an error rate in genotyping for these two chromosomes of 0.008 that is, in general, in agreement with the genome-wide figure quoted above. Extrapolating to the whole genome, if ~50% of the observed aberrant double recombinants represent genotyping errors, then the overall error rate is ~0.005 or 1 in 200 genotypes in the database. What is notable is that a significant number of 1(.)2(.)1 and 2(.)1(.)2 haplotypes remain on each chromosome despite this second round of error checking.

By and large, these aberrant haplotypes do not result from misorders because the level of triple recombinants is very low (see below). We have also considered the possibility that some or all of these aberrant double recombinants arise because of residual heterozygosity within the *M. spretus* mice used in establishing the backcross. That part of the EUCIB backcross performed in London used *M. spre-*

tus animals from a colony that had not been systematically inbred, whereas *M. spretus* animals used in Paris were from the SEG/Pas colony that is moderately inbred after 20 generations of unrelaxed brother-sister matings (Breen et al. 1994). Apparent double recombinant chromosomes of the form SSBSS (where S and B are the *M. spretus* and BL/6 alleles, respectively) could arise if there is residual heterozygosity in the *M. spretus* parents with a B rather than an S allele present at the supposed double recombinant locus in some members of the parent *M. spretus* population. In the backcross to BL/6, an SSBSS haplotype inherited from the F1 would be scored as a 2(.)1(.)2 haplotype in MBx—which we have designated a B1 haplotype. Conversely, in the backcross to *M. spretus*, an SSBSS haplotype would be scored as 1(.)2(.)1 in MBx, designated a S2 haplotype. Both B1 and S2 classes might be expected to be in excess if residual heterozygosity was a significant factor. However, residual heterozygosity from the *M. spretus* parent population could not account for haplotypes of the form BBSBB. In the backcross to BL/6, an BBSBB haplotype inherited from the F1 would be scored as a 1(.)2(.)1 haplotype in MBx—which we have designated a B2 haplotype. Conversely, in the backcross to *M. spretus*, an BBSBB haplotype would be scored as 2(.)1(.)2 in MBx—designated a S1 haplotype. Overall, we find that there are in total 606 double recombinant haplotypes in the B1 + S2 class, whereas there are 552 haplotypes in the B2 + S1 class. Residual heterozygosity does not, therefore, appear to be a major factor in the appearance of aberrant double recombinants.

Marker Order Error Rate

Following retyping and reordering, we also identified on every chromosome all triple recombinant haplotypes that remained—11(.)2(.)1(.)22 and 22(.)1(.)2(.)11—and that potentially represent local misorders. The numbers of triple recombinant haplotypes remaining on each chromosome are also given in Table 4. Some chromosomes had no detectable triple recombinants and in total, across the genome, we found 160 haplotypes representing the likely total number of locally misordered markers.

DISCUSSION

We have constructed the first high-resolution genetic map for a mammalian species. The EUCIB high-resolution microsatellite map has allowed us to order markers to 2302 bins providing a bin sepa-

Table 5. Assessing Genotype Error Rates in EUCIB

Chromosome	Double recombinants 1(.)2(.)1/2(.)1(.)2		
1	125 ^a	86 ^b	83 ^c
5	157 ^a	84 ^b	65 ^c

Reduction in the numbers of double recombinants following further rounds of re-genotyping and reordering on chromosomes 1 and 5 are given.

^aAfter data production, error checking, and ordering.

^bRemaining double recombinants following re-genotyping.

^cRemaining double recombinants following reorder.

ration at ~ 0.6 cM and a robust framework on which to complete the physical map of the mouse genome. Seventy-one percent of bins contain only one marker and only two large bins of 11 and 12 markers remain among the 3368 microsatellites mapped. Five gaps in the genetic map >5 cM remain on chromosomes 1, 5, 9, and 18. These do not correspond to regions of poor coverage on the EUCIB map because many of the Whitehead/MIT microsatellites from these regions were added to the map. In addition, the gaps in the EUCIB maps do not correspond to any of the larger genetic intervals remaining on the Whitehead/MIT genetic map (Dietrich et al. 1996). Overall, it would appear that these gaps potentially correspond to recombination hotspots on the EUCIB map.

Error checking confirmed the robustness of the genetic maps constructed. Most importantly, the total number of locally misordered markers identified by aberrant triple recombinants (see Results) was small—only 160 in the total dataset. To empirically estimate the genome-wide genotyping error rate remaining in the EUCIB dataset, we reanalyzed for chromosomes 1 and 5 all double recombinant 1(.)2(.)1 and 2(.)1(.)2 haplotypes for adjacent markers from within the dataset (see Results and Table 5). Approximately 50% of the original aberrant haplotypes were found to be incorrect by repeating the aberrant genotypings. Extrapolation to the whole genome gives an error rate for genotyping of ~ 0.005 . Given the low numbers of aberrant triple recombinants and the good agreement between the EUCIB and Whitehead/MIT maps, it would appear that a genotyping error rate of ~ 0.005 or lower will enable a reliable construction of high-resolution ordered maps from mouse backcrosses.

We were unable to eliminate a significant number of the aberrant double recombinant haplotypes of adjacent markers. As indicated above, the low numbers of aberrant triple recombinants and the general agreement of the EUCIB and Whitehead/MIT maps, gives us confidence that they do not all arise from local misordering. In addition, residual heterozygosity in the *M. spretus* parent population does not seem to be a major factor in the appearance of aberrant double recombinants (see Results). However, retesting of aberrant genotypes was carried out without altering the basic PCR reaction conditions. Under these circumstances, it remains to be determined by more extensive examination of the mapped remaining double recombinants whether they have failed to be eliminated because of technical reasons, or truly represent aberrant nonmendelian events.

The EUCIB resource will be used in two ways. Firstly, the provision of definitive order for many microsatellite markers at high resolution across the genome will assist in the verification of physical maps as well as aid in orientating contigs and contig closure. Secondly, EUCIB can be used to provide further resolution and robustness to the construction of genetic and physical maps in any chromosome region. It is important to recognize that the total number of recombinants in the EUCIB resource—17,029—far exceeds the minimum of 2302 recombinants required to separate and order the 2302 bins that form the EUCIB high-resolution map. Thus, the resolution of the EUCIB resource is not exhausted and clearly further recombinants can be exploited in any and every chromosome region to increase the resolution of genetic maps and to further enhance the construction of physical maps.

For a number of species (pig, cow, rat, and zebrafish) a number of genetic approaches with a variety of marker types are currently underway to develop comprehensive genome-wide microsatellite maps at intermediate resolution (Barendse et al. 1994; Postlethwait et al. 1994; Archibald et al. 1995; Jacob et al. 1995; Knapik et al. 1996). Local high-resolution genetic maps of STSs that have been constructed, either in the mouse or in other species, have proved enormously helpful in the construction of regional but robust physical maps. As for the mouse, there will be considerable value in ultimately developing genome-wide high-resolution maps in these other species.

METHODS

The EUCIB Backcross and MBx Database

The European Collaborative Interspecific Backcross (EUCIB) resource comprises 982 DNAs derived from a C57BL/6-*M. spretus* backcross and has already been described in detail (Breen et al. 1994). The MBx database, URL <http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html>, which holds all mouse, marker, and genotyping data and computes and displays high resolution maps has also been described (Breen et al. 1994). MBx can abstract and list all mice carrying single, double, and triple recombinant chromosomes (and chromosomes carrying larger numbers of apparent recombination events) across any chromosome region. In addition, MBx is able to provide a summary of haplotypes and their frequencies for any chromosome region. For access to the primary database contact: support@hgmp.mrc.ac.uk.

Details of map functions for the construction of genetic maps (see text) can be found at the MBx web site. Briefly, anchors and microsatellites are ordered on any chromosome using an algorithm that minimizes recombinants. This order is strictly maintained in constructing the genetic maps. The centromeric primary anchor on each chromosome is assigned a genetic map position according to consensus map data. In-

teranchor genetic distances for the primary anchors are then calculated for each chromosome and the remaining primary anchors assigned to the genetic map. Secondary anchors and microsatellites are subsequently incorporated into the genetic map maintaining genetic order as derived from the haplotype analysis. For each primary anchor interval, the total cumulative number of recombinants separating anchor and microsatellite markers in that interval is derived. The genetic distance separating a marker from any other marker or anchor in each primary anchor interval can then be calculated from the primary anchor genetic distance on the basis of the following ratio:

No. of recombinants separating marker from adjacent marker or anchor/Total cumulative no. of recombinants separating anchors and markers in the interval

Microsatellites mapping beyond the most proximal or distal primary anchors according to the haplotype analysis are added to the genetic map separately. However, it is likely that only a fraction of the relevant recombinants separating these markers and the primary anchor have been tested and genetic distances determined are, therefore, inaccurate.

Genotyping

Recombinant panels in any chromosome region were genotyped by use of a high-throughput, semiautomated fluorescent dUTP genotyping approach that has been described recently (Rhodes et al. 1997).

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council, UK and partly by the European Commission (grant no. GENE-CT-93-0046). We thank the EUCIB steering committee for their advice and support and comments on earlier drafts of this paper. We also thank David Papworth for statistical advice.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Andersson, L., A. Archibald, M. Ashburner, S. Audun, W. Barendse, J. Bitgood, C. Bottema, T. Broad, and S.D.M. Brown. 1996. Comparative genome organization of vertebrates (First International Workshop). *Mamm. Genome* 7: 717-734.
- Archibald, A.L., C.S. Haley, J.F. Brown, S. Couperwhite, H.A. McQueen, D. Nicholson, W. Coppieters, A. Van der Weghe, A. Stratil, A.-K. Wintero et al. 1995. The PiGMap consortium linkage map of the pig (*Sus scrofa*). *Mamm. Genome* 6: 157-175.
- Avner, P., L. Amar, L. Dandolo, and J.L. Guenet. 1988. Genetic analysis of the mouse using interspecific crosses. *Trends Genet.* 4: 18-23.
- Barendse, W., S.M. Armitage, L.M. Kossarek, A. Shalom, B.W. Kirkpatrick, A.M. Ryan, D. Clayton, L. Li, H.L. Neibergs, N. Zhang et al. 1994. A genetic linkage map of the bovine genome. *Nature Genet.* 6: 227-235.
- Breen, M., L. Deakin, B. Macdonald, S. Miller, R. Sibson, E. Tarttelin, P. Avner, F. Bourgade, J.-L. Guenet, X. Montagutelli et al. 1994. Towards high resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* 3: 621-627.
- Brown, S.D.M. 1994. Integrating maps of the mouse genome. *Curr. Opin. Genet. Dev.* 4: 389-394.
- . 1996. Mouse genome. In *Encyclopedia of molecular biology and molecular medicine* (ed. R.A. Meyers), Vol. 4, pp. 120-128. VCH Publishers, New York, NY.
- Brown, S.D.M. and J. Peters. 1996. Combining mutagenesis and genomics in the mouse—closing the phenotype gap. *Trends Genet.* 12: 433-435.
- Ceci, J.D., L.D. Siracusa, N.A. Jenkins, and N.G. Copeland. 1989. A molecular genetic linkage map of mouse chromosome 4 including the localization of several proto-oncogenes. *Genomics* 5: 699-709.
- Copeland, N. and N.A. Jenkins. 1991. Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* 7: 113-118.
- Copeland, N.G., N.A. Jenkins, D.J. Gilbert, J.T. Eppig, L.J. Maltais, J.C. Miller, W.F. Dietrich, A. Weaver, S.E. Lincoln, R.G. Steen et al. 1993. A genetic linkage map of the mouse: Current applications and future prospects. *Science* 262: 57-66.
- Dietrich, W.F., H. Katz, S.E. Lincoln, H.S. Shin, J. Friedman, N.C. Dracopoli, and E.S. Lander. 1992. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131: 423-447.
- Dietrich, W.F., N.G. Copeland, D.J. Gilbert, J.C. Miller, N.A. Jenkins, and E.S. Lander. 1995. Mapping the mouse genome: Current status and future prospects. *Proc. Natl. Acad. Sci.* 92: 10849-10853.
- Dietrich, W.F. et al. 1996. A comprehensive genetic map of the mouse genome. *Nature* 380: 149-152.
- Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.-H. Xu et al. 1996. An STS-based map of the human genome. *Science* 270: 1945-1954.
- Jacob, H.J., D.M. Brown, R.K. Bunker, M.J. Daly, V.J. Dzau, A. Goodman, G. Koike, V. Kren, T. Kurtz, A. Lernmark et al. 1995. Genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nature Genet.* 9: 63-69.
- Knapik, E.W., A. Goodman, O. Scott Atkinson, C.T. Roberts, M. Shiozawa, C.U. Sim, S. Weksler-Zangen, M.R. Trolliet, C. Futrell, B.A. Innes et al. 1996. A reference cross DNA panel for zebrafish (*Danio rerio*) anchored with simple sequence length polymorphisms. *Development* 123: 451-460.

RHODES ET AL.

Nadeau, J.H., B. Herrman, M. Bucan, D. Burkart, J.L. Crosby, M.A. Erhart, M. Kosowsky, J.P. Kraus, F. Michiels, A. Schnattinger et al. 1991. Genetic maps of mouse chromosome 17 including 12 new anonymous DNA loci and 25 anchor loci. *Genomics* 9: 78-89.

Postlethwait, J.H., S.L. Johnson, C.N. Midson, W.S. Talbot, M. Gates, E.W. Ballinger, D. Africa, R. Andrews, T. Carl, J.S. Eisen et al. 1994. A genetic linkage map for the zebrafish. *Science* 264: 699-703.

Reeves, R.H., M.R. Crowley, W.S. Moseley, and M.F. Seldin. 1991. Comparison of interspecific to intersubspecific backcrosses demonstrates species and sex differences in recombination frequency on mouse Chromosome 16. *Mamm. Genome* 1: 158-164.

Reeves, R.H., E.E. Rue, M.P. Citron, and D.E. Cabin. 1997. High-resolution recombinational map of mouse chromosome 16. *Genomics* 43: 202-208.

Rhodes, M., A. Dearlove, R. Straw, S. Fernando, A. Evans, M. Greener, T. Lacey, M. Kelly, K. Gibson, S.D.M. Brown, and C. Mundy. 1996. High throughput microsatellite analysis using fluorescent dUTPs for high resolution genetic mapping of the mouse genome. *Genome Res.* 7: 81-86.

Saunders, A.M. and M.F. Seldin. 1990. A molecular genetic linkage map of mouse chromosome 7. *Genomics* 8: 525-535.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, R. Rodriguez-Tome, A. Aggarwal, E. Bajorek et al. 1996. A gene map of the human genome. *Science* 274: 540-546.

Vignal, A., G. Gyapay, J. Hazan, S. Nguyen, C. Dupraz, N. Cheron, N. Becuwe, M. Tranchant, and J. Weissenbach. 1993. Non radioactive Multiplex Procedure for Genotyping of Microsatellite Markers. In *Methods in molecular genetics* ed. Kenneth W. Adolph, Vol. 1, pp. 211-221. Academic Press, San Diego, CA.

Received August 4, 1997; accepted in revised form February 9, 1998.

Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes

(polymerase chain reaction/"reverse dot blots"/nonradioactive detection/*HLA-DQA* locus/ β -thalassemia)

RANDALL K. SAIKI*, P. SEAN WALSH*, COREY H. LEVENSON†, AND HENRY A. ERLICH*

Departments of *Human Genetics and †Chemistry, Cetus Corp., 1400 Fifty-Third Street, Emeryville, CA 94608

Communicated by Hamilton O. Smith, May 9, 1989 (received for review March 2, 1989)

ABSTRACT The analysis of DNA for the presence of particular mutations or polymorphisms can be readily accomplished by differential hybridization with sequence-specific oligonucleotide probes. The *in vitro* DNA amplification technique, the polymerase chain reaction (PCR), has facilitated the use of these probes by greatly increasing the number of copies of target DNA in the sample prior to hybridization. In a conventional assay with immobilized PCR product and labeled oligonucleotide probes, each probe requires a separate hybridization. Here we describe a method by which one can simultaneously screen a sample for all known allelic variants at an amplified locus. In this format, the oligonucleotides are given homopolymer tails with terminal deoxyribonucleotidyltransferase, spotted onto a nylon membrane, and covalently bound by UV irradiation. Due to their long length, the tails are preferentially bound to the nylon, leaving the oligonucleotide probe free to hybridize. The target segment of the DNA sample to be tested is PCR-amplified with biotinylated primers and then hybridized to the membrane containing the immobilized oligonucleotides under stringent conditions. Hybridization is detected nonradioactively by binding of streptavidin-horseradish peroxidase to the biotinylated DNA, followed by a simple colorimetric reaction. This technique has been applied to *HLA-DQA* genotyping (six types) and to the detection of Mediterranean β -thalassemia mutations (nine alleles).

Differential hybridization with sequence-specific oligonucleotide probes has become a widely used technique for the detection of genetic mutations and polymorphisms (1-5). When hybridized under the appropriate conditions, these synthetic DNA probes (usually 15-20 bases in length) will anneal to their complementary target sequences in the sample DNA only if they are perfectly matched. In most cases, the destabilizing effect of a single base-pair mismatch is sufficient to prevent the formation of a stable probe-target duplex (6). With an appropriate selection of oligonucleotide probes, the relevant genetic content of a DNA sample can be completely described.

This very powerful method of DNA analysis has been greatly simplified by the *in vitro* DNA-amplification technique, the polymerase chain reaction (PCR) (7-9). The PCR can selectively increase the number of copies of a particular DNA segment in a sample by many orders of magnitude. As a result of this 10^6 - to 10^8 -fold amplification, more convenient assays and nonradioactive detection methods have become possible (10-12). These PCR-based assays are usually done by amplifying the target segment in the sample to be tested, fixing the amplified DNA onto a series of nylon membranes, and hybridizing each membrane with one of the labeled oligonucleotide probes under stringent hybridization conditions. However, each probe must still be individually hybrid-

ized to the amplified DNA and the process can easily become difficult in a system where many different mutations or polymorphisms occur.

One approach to address this procedural difficulty is to "reverse" the DNAs: attach the oligonucleotides to the nylon support and hybridize the amplified sample to the membrane. Thus, in a single hybridization reaction, an entire series of sequences could be analyzed simultaneously. The strategy we adopted was to immobilize the oligonucleotides onto nylon filters by ultraviolet fixation. Exposure to UV light activates thymine bases in DNA, which then covalently couple to the primary amines present in nylon (13). It seemed unlikely, however, that short oligonucleotides could be directly attached to nylon in this manner and still retain their ability to discriminate at the level of a single base-pair mismatch. Consequently, the addition of a long deoxyribothymidine homopolymer tail, poly(dT), to the 3' end of the oligonucleotide appeared promising for several reasons. First, the poly(dT) tail would be a larger target for UV crosslinking and should preferentially react with the nylon. Second, dTTP is very readily incorporated onto the 3' ends of oligonucleotides by terminal deoxyribonucleotidyltransferase and would permit the synthesis of very long tails (14). (Deoxyribothymidine would also be the most efficiently incorporated base if a purely synthetic route were chosen.) Third, Collins and Hunsaker (15) had shown that the presence of a poly(dA) homopolymer tail, used to introduce multiple ^{35}S labels, did not affect the function of sequence-specific oligonucleotide probes.

We have used this technique to attach oligonucleotide probes specific for the six major *HLA-DQA* DNA types (16) and the eight most common Mediterranean β -thalassemia mutations (4) to nylon filters. The target segment of the DNA sample to be tested (either *HLA-DQA* or β -globin) was amplified by PCR with biotin-labeled primers to introduce a nonradioactive tag. Hybridization of the amplified product to the immobilized oligonucleotides and binding of streptavidin-horseradish peroxidase conjugate to the biotinylated primers were performed simultaneously. Detection was accomplished by a simple colorimetric reaction involving the enzymatic oxidation of a colorless chromogen that yielded a red color wherever hybridization occurred.

MATERIALS AND METHODS

Tailing of Oligonucleotides. Oligonucleotides were synthesized on a DNA synthesizer (model 8700, Bioscience) with β -cyanoethyl *N,N*-diisopropylphosphoramidite nucleosides (American Bionetics, Hayward, CA) by using protocols provided by the manufacturer. Oligonucleotide (200 pmol) was tailed in 100 μl of 100 mM potassium cacodylate/25 mM Tris-HCl/1 mM CoCl_2 /0.2 mM dithiothreitol, pH 7.6 (17), with 5-160 nmol deoxyribonucleoside triphosphate (dTTP or

dCTP) and 60 units (50 pmol) of terminal deoxyribonucleotidyltransferase (Ratloff Biochemicals, Los Alamos, NM) for 60 min at 37°C. Reactions were stopped by addition of 100 μ l of 10 mM EDTA. The lengths of the homopolymer tails were controlled by limiting dTTP or dCTP. For example, a nominal tail length of 400 dT residues was obtained by using 80 nmol of dTTP in the above reaction.

Preparation of Filters. The tailed oligonucleotides were diluted into 100 μ l of TE (10 mM Tris-HCl/0.1 mM EDTA, pH 8.0) and applied to a nylon membrane (Genetran-45; Plasco, Woburn, MA) with a spotting manifold (BioDot; BioRad). The damp filters were then placed on TE-soaked paper pads in a UV light box (Stratalinker 1800; Stratagene) and irradiated at 254 nm. Dosage was controlled by the device's internal metering unit. The irradiated membranes were washed in 200 ml of 5 \times SSPE (1 \times SSPE is 180 mM NaCl/10 mM NaH₂PO₄/1 mM EDTA, pH 7.2) with 0.5% NaDodSO₄ for 30 min at 55°C to remove unbound oligonucleotides. If not used immediately, the filters were rinsed in water, air-dried, and stored at room temperature until needed.

Amplification of DNA. PCR amplification of genomic sequences was performed by a slight modification of previously described procedures (9). DNA (0.1–0.5 μ g) was amplified in 100 μ l containing 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 1.5 mM MgCl₂, 10 μ g of gelatin, 200 μ M each dATP, dCTP, dGTP, and dTTP, 0.2 μ M each biotinylated amplification primer, and 2.5 units of *Thermus aquaticus* (Taq) DNA polymerase (Perkin-Elmer/Cetus). The cycling reaction was done in a programmable heat block (DNA Thermal Cycler; Perkin-Elmer/Cetus) set to heat at 95°C for 15 sec (denature), cool at 55°C for 15 sec (anneal), and incubate at 72°C for 30 sec (extend) by the "Step-Cycle" program. After 30 repetitions, the samples were incubated an additional 5 min at 72°C. The primers contained a single molecule of biotin attached to the 5' end of the oligonucleotides (described below).

Hybridization and Detection of Amplified DNA. Each filter with bound oligonucleotides was placed in 4 ml of hybridization solution containing 5 \times SSPE, 0.5% NaDodSO₄, and 400 ng of streptavidin-horseradish peroxidase conjugate (SeeQuence; Eastman Kodak). PCR-amplified DNA (20 μ l) was denatured by addition of an equal volume of 400 mM NaOH/10 mM EDTA and added immediately to the hybridization solution, which was then incubated at 55°C for 30 min. (During this incubation, hybridization of PCR product to immobilized oligonucleotide and binding of streptavidin-horseradish peroxidase to biotin present in the PCR product occur simultaneously.) The filters were briefly rinsed twice in 1 \times SSPE/0.1% NaDodSO₄ at room temperature, washed once in 2 \times SSPE/0.1% NaDodSO₄ at 55°C for 10 min, and then briefly rinsed twice in 2 \times PBS (1 \times PBS is 137 mM NaCl/2.7 mM KCl/8 mM Na₂HPO₄/1.5 mM KH₂PO₄, pH 7.4) at room temperature. Color development was performed by incubating the filters in 25–50 ml of red leuco dye (Eastman Kodak) at room temperature for 5–10 min. Photographs were taken for permanent records.

Synthesis of Biotinylated Oligonucleotide Primers. Primary amino groups were introduced at the 5' termini of the primers by a variation of published procedures (18, 19). In brief, ethylene glycol was converted to the monophthalimido derivative by reaction with phthalimide in the presence of triphenylphosphine and diisopropyl azodicarboxylate (20). The monophthalimide was converted to the corresponding β -cyanoethyl diisopropylamino phosphoramidite by standard protocols (21). The resulting phthalimido amidite was added to the 5' ends of the oligonucleotides during the final cycle of automated DNA synthesis by using standard coupling conditions. During normal deprotection of the DNA (concentrated aqueous ammonia for 5 hr at 55°C), the phthalimido

group was acylated with an appropriate biotin active ester. NHS-LC-biotin (Pierce) was selected for its water solubility and lack of steric hindrance. The biotinylation was performed on crude, deprotected oligonucleotide, and the mixture was purified by a combination of gel filtration and reversed-phase HPLC. Additional details of this procedure will be published elsewhere (22).

RESULTS

Binding and Hybridization Efficiency of Tailed Oligonucleotides. The relative efficiencies with which synthetic oligonucleotides with homopolymer tails of various lengths were covalently bound to the nylon filter were measured as a function of UV exposure (Fig. 1 *Left*). Oligonucleotides with longer poly(dT) tails were more readily fixed to the membrane, and all attained their maximum values by 240 mJ/cm² of irradiation at 254 nm. In contrast, the (dC)₄₀₀-tailed oligonucleotide required more irradiation to crosslink to the nylon and was not comparable to the equivalent (dT)₄₀₀ construct even after 600 mJ/cm² exposure. This difference is consistent with the findings of Church and Gilbert (13) that suggested light-activated thymine bases bind more effectively to nylon than do cytosine bases. The untailed oligonucleotide was also retained by the membrane in a manner that roughly paralleled the poly(dC) product.

Efficient binding of oligonucleotides to the membrane, however, does not necessarily correlate with hybridization efficiency, and so hybridization efficiency as a function of UV dosage was determined in a separate experiment (Fig. 1 *Right*). These results show a distinct optimum of exposure that changes with the length of the poly(dT) tail and is more sharply pronounced for the longer tails. Additional experiments have shown the optimal dosages to be about 20 mJ/cm² for the (dT)₈₀₀ and 40 mJ/cm² for the (dT)₄₀₀ oligonucleotides (R.K.S., unpublished observations). The peak efficiencies of the (dT)₄₀₀ and (dT)₈₀₀ constructs are around 1% (45–50 fmol of radiolabeled probe annealed to \approx 3.5 pmol of tailed oligonucleotide), which is similar to the value reported by Gamper *et al.* (23) for an oligonucleotide probe hybridized to nylon-bound plasmid DNA.

Comparison of the data in Fig. 1 *Left* and *Right* for 60 mJ/cm² irradiation indicates that oligonucleotides with longer tails hybridize more effectively than can be accounted for by the additional amounts bound to the filter. This suggests a spacer effect wherein the poly(dT) tails improve hybridization efficiency by increasing the distance between the nylon membrane and the terminal oligonucleotide probe. Besides possible UV damage to the DNA itself, additional exposure causes more of the tail to become attached to the membrane, thus reducing the average spacer length and decreasing hybridization efficiency. The markedly different hybridization profile of the poly(dC) oligonucleotide is compatible with this interpretation. Because cytosines react less efficiently with the filter, hybridization efficiency reaches a plateau where loss due to UV damage and tail shortening are compensated by the fixing of new molecules (see Fig. 1 *Left*). This characteristic of cytosine may make a poly(dC) tail desirable when UV irradiation cannot be carefully controlled. Under the stringent hybridization conditions used in this experiment, no signal was detected for the untailed oligonucleotide.

DNA Typing at the HLA-DQA Locus. The HLA-DQA test is derived from a PCR-based oligonucleotide typing system that partitions the polymorphic variants at the DQA locus into four major DNA types, DQA1 to DQA4, and three DQA1 subtypes, DQA1.1 to DQA1.3 (16). Four oligonucleotides specific for the major DQA types, four oligonucleotides that characterize the DQA1 subtypes, and one control oligonu-

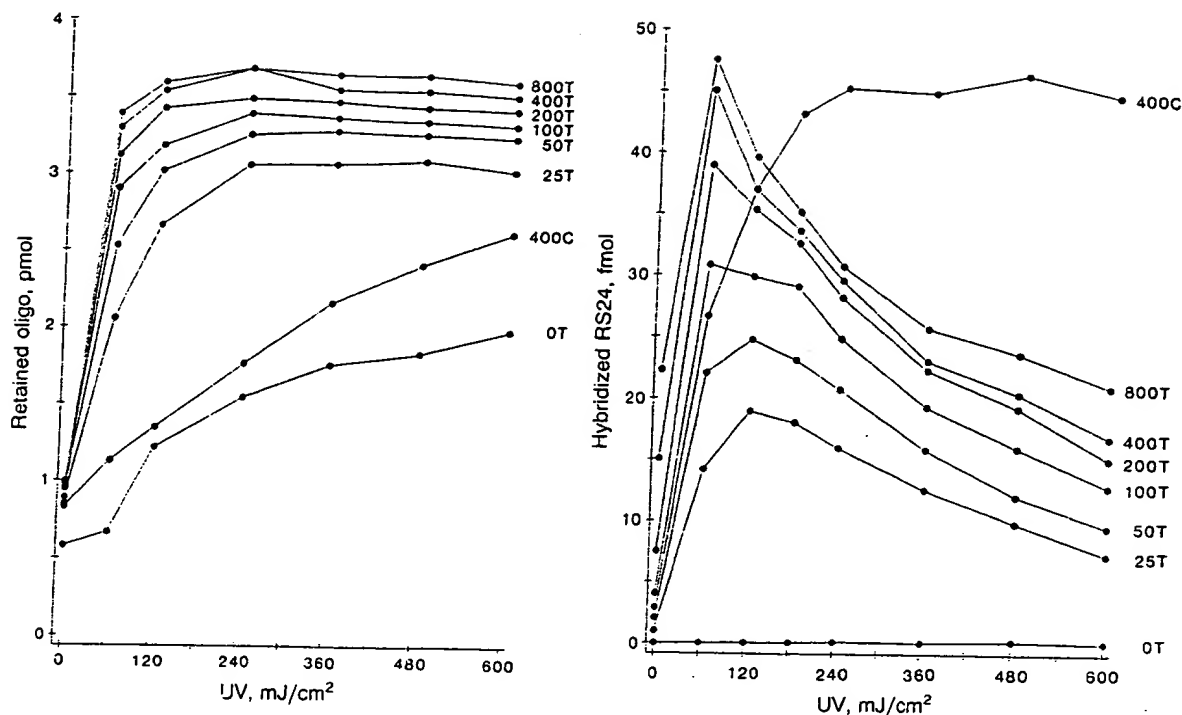


FIG. 1. Filter retention and hybridization efficiency of tailed oligonucleotides as a function of UV dosage and tail length. (Left) Filter retention. A 19-base oligonucleotide, 19A (5'-CTCCTGAGGAGAAGTCTGC-3'), was 5'-end-labeled with ^{32}P by using phage T4 polynucleotide kinase and [$\gamma\text{-}^{32}\text{P}$]ATP (10). Portions of the labeled oligonucleotide were given 3' homopolymer tails with terminal deoxynucleotidyltransferase and either dTTP or dCTP. The base compositions and lengths of the tails were as follows: (dT)₀, (dT)₂₅, (dT)₅₀, (dT)₁₀₀, (dT)₂₀₀, (dT)₄₀₀, (dT)₈₀₀, and (dC)₄₀₀. Four picomoles of each oligonucleotide was spotted onto nine duplicate filters, UV irradiated for various times, and washed to remove unbound oligonucleotides; each spot then was measured by scintillation counting to determine the amount crosslinked to the nylon. The values plotted are relative to an unirradiated, unwashed control filter (100% retention). (Right) Hybridization efficiency. Filters containing tailed, but unlabeled, 19A were prepared as described above and hybridized under sequence-specific conditions (see Materials and Methods) with a ^{32}P -labeled 40-base oligonucleotide, RS24 (5'-CCCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAG-3'), complementary to 19A. The specific activity of the RS24 was 1.5 $\mu\text{Ci}/\text{pmol}$ (1 μCi = 37 kBq). Each spot was assayed by scintillation counting. The values plotted are fmol of RS24 hybridized to the membrane.

1) were given 400-base poly(dT) tails and spotted onto nylon filters. The sequence variation that defines the *DQA* types is localized within a relatively small "hypervariable" region of the second exon (24) that can be encompassed within a single 242-base-pair PCR amplification fragment. Biotinylated primers (Table 1) were used to amplify the *DQA* fragment from several genomic DNA samples: six homozygous cell lines and six heterozygous individuals. After hybridization of the amplified DNA to the membranes and color development, the *DQA* genotypes of these samples were readily apparent (Fig. 2).

Although most of the oligonucleotide probes are unique, specific for one *DQA* type, two of the *DQA1* subtyping probes cross-hybridize to several DNA types. GH89 hybridizes to a sequence common to the *DQA1.2*, *DQA1.3*, and *DQA4* types, and the probe GH76 detects all *DQA* types except *DQA1.3*. (The latter is needed to distinguish *DQA1.2/1.3* heterozygotes from *DQA1.3/1.3* homozygotes.) The length and strand specificity of the oligonucleotides were empirically adjusted until their relative hybridization efficiencies and stringency requirements for allelic discrimination were approximately the same. (This was achieved by deter-

Table 1. Sequences of oligonucleotide primers and probes

Name*	Function	Sequence	Name*	Function	Sequence
RS151	<i>DQA</i> primer	b-GTGCTGCAGGTGTAACTTGTACCAG [†]	RS151	β -Globin primer	b-ATCACTTAGACCTCACCCCTG [†]
RS152	<i>DQA</i> primer	b-CACGGATCCGGTAGCAGCGGTAGAGTTG [†]	RS152	β -Globin primer	b-GACCTCCCACATTCCCTTTT [†]
RH54 (2)	All <i>DQA</i> types	CTACGTGGACCTGGAGAGGAAGGAGACTGCCTG	RS187 (8)	Normal β^{1-110}	TAGACCAATAGGCAGAGAG
GH75 (4)	<i>DQA1</i> probe	CTCAGGCCACCGCCAGGCA	RS188 (8)	Mutant β^{1-110}	CTCTCTGCTATTAGTCTA
RH71 (4)	<i>DQA2</i> probe	TTCCACAGACTTAGATTGAC	RS87 (4)	Normal β^{39}	CCTTGGACCCAGAGGTTCT
GH67 (4)	<i>DQA3</i> probe	TTCCGCAGATTAGAAGAT	RS89 (4)	Mutant β^{39}	AGAACCCTCTAGGTCCAAGG
GH66 (4)	<i>DQA4</i> probe	TGTTTGCCTGTTCTCAGAC	RS189 (0.33)	Normal $\beta^{1-1.6}$	CTTGATACCAACCTGCCCA [‡]
GH88 (4)	<i>DQA1.1</i> probe	CGTAGAACTCCTCATCTCC	RS190 (0.33)	Mutant β^{1-6}	TGGGCAGGTTGGCATCAAG
GH89 (4)	<i>DQA1.2, -1.3, -4</i>	GATGAGCAGTTCTACGTGG	RS191 (1)	Mutant β^{1-1}	TGGGCAGATTGGTATCAAG
GH77 (4)	<i>DQA1.3</i> probe	CTGGAGAAGAAGGAGAC	RS192 (4)	Normal β^{2-1}	CCATAGACTCACCCCTGAAG
GH76 (4)	Not <i>DQA1.3</i>	GTCTCTCTCTCCAG	RS193 (4)	Mutant β^{2-1}	CTTCAGGATGAGTCTATGG
			RS201 (2)	Normal β^{2-745}	GCAGAATGGTAGCTGGATT
			RS202 (2)	Mutant β^{2-745}	GCAGAATGGTACCTGGATT
			RS196 (4)	Normal $\beta^{6,8}$	ACTCCTGAGGAGAAGTCTG [‡]
			RS197 (4)	Mutant β^6	GACTCCTGGGAGAAGTCTG
			RS198 (4)	Mutant β^8	TGACTCCTGAGGAGGTCTG

*Where applicable, the values in parentheses indicate the amount (pmol) of tailed oligonucleotide probe applied to the nylon membrane.

[†]b, Biotin covalently attached to 5' end.

[‡]These β -globin oligonucleotide probes each span two sites of potential β -thalassemia mutations and are specific for normal sequences at both positions.

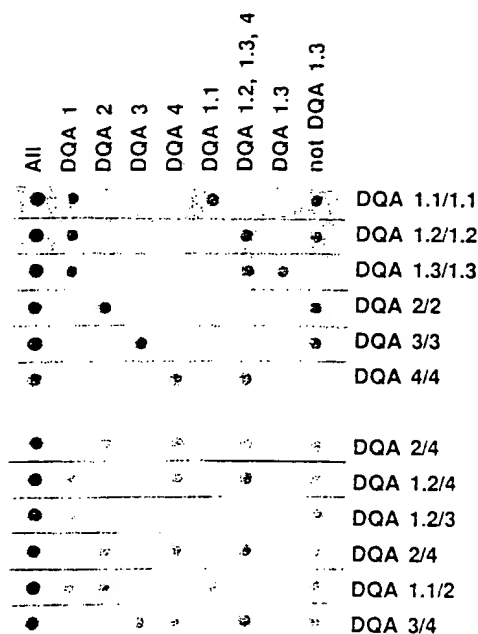


FIG. 2. DNA typing at the *HLA-DQA* locus. Each tailed oligonucleotide probe was spotted onto 12 duplicate membranes, irradiated at 40 mJ/cm², hybridized with amplified *DQA* sequences in genomic DNA samples, and treated for color development. The specificity of each immobilized oligonucleotide is given at the top, and the *DQA* genotype of each sample is noted at the right. The name, amount applied to the membrane, specificity, and sequence of each oligonucleotide are listed in Table 1.

mining the optimal hybridization conditions for each member of an initial set of probes, then shortening or lengthening each oligonucleotide until they all hybridized under equivalent conditions.) These eight probes produce a unique hybridization pattern for each of the 21 possible *DQA* diploid combinations.

Detection of β -Thalassemia Mutations. Although there are >54 characterized mutations of the β -globin gene that can give rise to β -thalassemia (25), each ethnic group in which this disease is prevalent has a limited number of common mutations (4, 26, 27). In Mediterranean populations, 8 mutations are responsible for >90% of the β -thalassemia alleles (4). Oligonucleotides were synthesized that are specific for each of these 8 mutations as well as their corresponding normal sequences (Table 1). The oligonucleotides were given

(dT)₄₀₀ tails with terminal transferase and applied to membranes. Since the β -thalassemia mutations are distributed throughout the β -globin gene, biotinylated PCR primers that amplify the entire gene in a single 1780-base-pair fragment were used. (This amplification product encompasses all known β -thalassemia mutations, not only the predominant Mediterranean mutations examined here.) After hybridization and color development, the β -globin genotypes could be determined by noting the pattern of hybridization (Fig. 3).

Unlike the *DQA* typing system, two oligonucleotide probes are needed to analyze each mutation—one specific for the normal sequence and one specific for the mutant sequence—in order to differentiate normal/mutant heterozygous carriers from mutant/mutant homozygotes. A complicating factor in this analysis is caused by apparent secondary structure in various portions of the relatively long β -globin amplification product that interferes with oligonucleotide hybridization. The relatively high stringency needed to minimize this secondary structure requires the use of longer (e.g., 19-base) oligonucleotide probes. Because this constraint would not permit varying the length of the oligonucleotides to compensate for different hybridization efficiencies, the "balancing" of signal intensities was accomplished by adjusting the amount of each oligonucleotide applied to the membrane. This was done by applying various amounts of each oligonucleotide onto a membrane and then, after hybridization and color development, simply selecting the positive spots that had similar intensity.

DISCUSSION

These studies have demonstrated the feasibility of immobilizing sequence-specific probes onto nylon membranes and hybridizing PCR-amplified, biotin-labeled genomic fragments to the filters to determine the genetic content of the DNA sample. We have applied this method to *HLA-DQA* genotyping and to the detection of β -thalassemia mutations. Although the number of probes used in the two tests were modest (9 for *DQA* and 14 for β -thalassemia), expanding the analyses to include even more oligonucleotides should not be difficult.

The recently described technique of simultaneous amplification of several DNA fragments, "multiplex" PCR (28), should readily permit the concurrent analysis of multiple genetic loci. Using the immobilized-probe format, we have been able to simultaneously amplify and type at three loci: the *Hind*III polymorphism in the α -globin gene (29), the *Ava* II

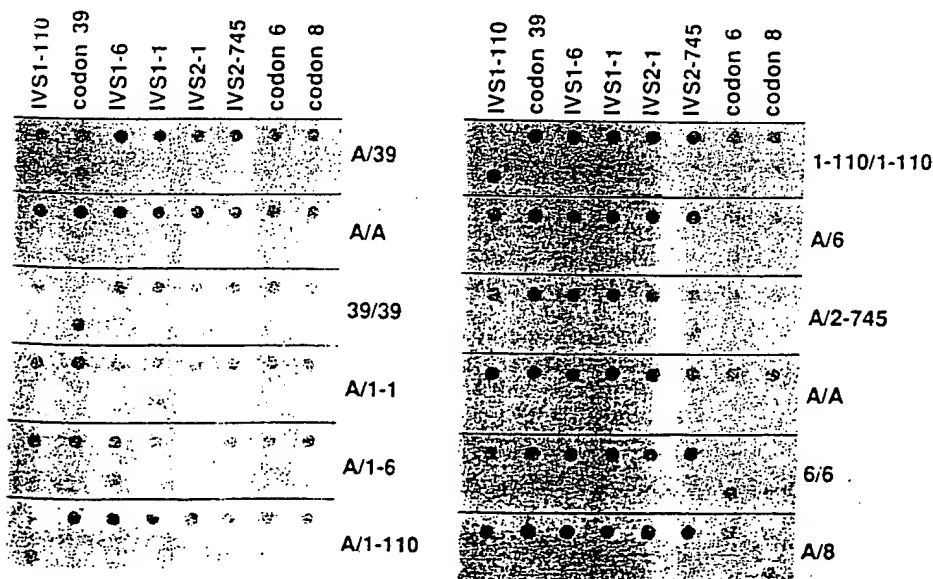


FIG. 3. Detection of β -thalassemia mutations. Various amounts of each tailed oligonucleotide probe were applied to 12 duplicate nylon filters, irradiated at 40 mJ/cm², hybridized with amplified β -globin sequences in genomic DNA samples, and treated for color development. The β -thalassemia locus that is detected by each immobilized oligonucleotide pair is given at the top of the filters. For each filter, the upper row contains the oligonucleotide probes that are specific for the normal sequence and the lower row contains the oligonucleotides specific for the mutant sequences. The β -globin genotype of each sample is noted at the right. The name, amount applied to the membrane, specificity, and sequence of each oligonucleotide are listed in Table 1. IVS, intervening sequence (intron).

polymorphism in the low density lipoprotein receptor gene (30), and the *HLA-DQA* gene (R.K.S., unpublished observations). Other genetic targets whose analysis would be simplified by this technique include the detection of somatic mutations in the *RAS* genes, where 6 loci and 66 possible alleles occur (31), some of the HLA class II β -chain genes, where as many as 25 alleles can be detected (T. Bugawan, S. Scharf, and H.A.E., unpublished observations), and β -thalassemia in Middle Eastern populations, where in addition to the endogenous mutations, Mediterranean and Asian Indian mutations are present at significant frequencies (H. Kazazian, personal communication). This format should also prove useful for the detection of infectious pathogens or for environmental surveys of microorganisms by immobilizing a panel of species-specific probes.

The ability to label probes and detect their hybridization without radioactivity is a convenient feature of PCR-based DNA tests and, perhaps more importantly, makes this type of analysis feasible in areas where radioactive labeling reagents are difficult to obtain. In this report, a biotin tag was introduced into the PCR products by means of 5'-biotinylated primers. An alternative labeling strategy based on the incorporation of biotinylated dUTP (32) has also been tried and shown to be very effective (R.K.S., unpublished observations).

One of the prerequisites of this analytical method is that all of the bound oligonucleotides must be sequence-specific under the same hybridization conditions. If necessary, this requirement can probably be met either by adjusting the length, position, and strand specificity of the probes, as was done for the *HLA-DQA* assay, or by varying the amount applied to the membrane, as was done for the β -thalassemia assay. The presence of tetramethylammonium chloride in the hybridization buffer can also serve to minimize the differences among immobilized oligonucleotides caused by varying base compositions (ref. 33; T. Bugawan, personal communication).

Although it may entail some initial effort, the end result is a simple, robust, and potentially automatable system that can be completed (amplification, hybridization, and color development) in 3–4 hr. "Reverse dot blots" should be particularly valuable for assays where the number of potential sequence variations exceeds the number of samples to be tested. Even in situations where the number of samples and probes are approximately equal, the immobilized-probe format may be preferable since many filters can be prepared at one time and stored until needed. To date, this typing system has been used to determine the *HLA-DQA* genotype of >300 unknown samples in forensic and disease-susceptibility studies.

We thank R. Higuchi and S. Scharf for helpful suggestions, L. Goda, D. Spasic, and C.-A. Chang for synthesis of oligonucleotides, C. Perez for advice on terminal transferase tailing reaction, C. Dowling and H. Kazazian (Johns Hopkins) for sequences of β -globin PCR primers and β -thalassemia genomic DNA samples, S. Warren and J. Findlay (Eastman Kodak) for red leuco dye suspension, and T. White, D. Gelfand, and H. Kazazian for critical review of the manuscript.

1. Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R. L., & Wallace, R. B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 278–282.
2. Pirastu, M., Kan, Y. W., Cao, A., Conner, B. J., Teplitz, R. L., & Wallace, R. B. (1983) *N. Engl. J. Med.* **309**, 284–287.
3. Bos, J. L., Verlaan-de Vries, M., Jansen, A. M., Veeneman, G. H., & van Boom, J. H. (1984) *Nucleic Acids Res.* **12**, 9155–9163.
4. Kazazian, H. H., Jr., Orkin, S. H., Markham, A. F., Chapman, C. R., Youssoufian, H., & Waber, P. G. (1984) *Nature (London)* **310**, 152–154.
5. Kidd, V. J., Golbus, M. S., Wallace, R. B., Itakura, K., & Woo, S. L. (1984) *N. Engl. J. Med.* **310**, 639–642.
6. Ikuta, S., Takagi, K., Wallace, R. B., & Itakura, K. (1987) *Nucleic Acids Res.* **15**, 797–811.
7. Mullis, K. B., & Faloona, F. A. (1987) *Methods Enzymol.* **155**, 335–350.
8. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985) *Science* **230**, 1350–1354.
9. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science* **239**, 487–491.
10. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1986) *Nature (London)* **324**, 163–166.
11. Bugawan, T. L., Saiki, R. K., Levenson, C. H., Watson, R. W., & Erlich, H. A. (1988) *Bio/Technology* **6**, 943–947.
12. Saiki, R. K., Chang, C.-A., Levenson, C. H., Warren, T. C., Boehm, C. D., Kazazian, H. H., Jr., & Erlich, H. A. (1988) *N. Engl. J. Med.* **319**, 537–541.
13. Church, G. M., & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
14. Nelson, T., & Brutlag, D. (1979) *Methods Enzymol.* **68**, 41–57.
15. Collins, M. L., & Hunsaker, W. R. (1985) *Anal. Biochem.* **141**, 211–224.
16. Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F., & Erlich, H. A. (1988) *Nature (London)* **332**, 543–546.
17. Roychoudhury, R., & Wu, R. (1980) *Methods Enzymol.* **65**, 43–62.
18. Coull, J. M., Weith, H. L., & Bischoff, R. (1986) *Tetrahedron Lett.* **27**, 3991–3994.
19. Connolly, B. A. (1987) *Nucleic Acids Res.* **15**, 3131–3139.
20. Mitsunobu, O. (1981) *Synthesis*, 1–28.
21. Sinha, N. D., McManus, J., & Koster, H. (1984) *Nucleic Acids Res.* **12**, 4539–4557.
22. Levenson, C. H., & Chang, C.-A. (1989) in *PCR Protocols and Applications: A Laboratory Manual*, eds. Innis, M. A., Gelfand, D. H., & White, T. J. (Academic, New York), in press.
23. Gamper, H. B., Cimino, G. D., Isaacs, S. T., Ferguson, M., & Hearst, J. E. (1986) *Nucleic Acids Res.* **14**, 9943–9954.
24. Horn, G. T., Bugawan, T. L., Long, C. M., & Erlich, H. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6012–6016.
25. Kazazian, H. H., Jr., & Boehm, C. D. (1988) *Blood* **72**, 1107–1116.
26. Kazazian, H. H., Jr., Orkin, S. H., Antonarakis, S. E., Sexton, J. P., Boehm, C. D., Goff, S. C., & Waber, P. G. (1984) *EMBO J.* **3**, 593–596.
27. Zhang, J. Z., Cai, S. P., He, X., Lin, H. X., Lin, H. J., Huang, Z. G., Chehab, F. F., & Kan, Y. W. (1988) *Hum. Genet.* **78**, 37–40.
28. Chamberlain, J. S., Gibbs, R. A., Ranier, J. E., Nguyen, P. N., & Caskey, C. T. (1988) *Nucleic Acids Res.* **16**, 11141–11156.
29. Jeffreys, A. J. (1979) *Cell* **18**, 1–10.
30. Hobbs, H. H., Esser, V., & Russell, D. W. (1987) *Nucleic Acids Res.* **15**, 379.
31. Verlaan-de Vries, M., Bogaard, M. E., van den Elst, H., van Boom, J. H., van der Eb, A. J., & Bos, J. L. (1986) *Gene* **50**, 313–320.
32. Lo, Y. M., Mehal, W. Z., & Fleming, K. A. (1988) *Nucleic Acids Res.* **16**, 8719–8720.
33. Wood, W. I., Gitschier, J., Lasky, L. A., & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585–1588.

A Primer-Guided Nucleotide Incorporation Assay in the Genotyping of Apolipoprotein E

ANN-CHRISTINE SYVÄNEN, KATRIINA AALTO-SETÄLÄ,* LEENA HARJU
KIMMO KONTULA,* AND HANS SÖDERLUND

Orion Corporation, Orion Pharmaceutica, Biotechnology, Valimotie 7, 00380 Helsinki, Finland; and
*Institute of Biotechnology, University of Helsinki, Valimotie 7, 00380 Helsinki, Finland

Received April 9, 1990; revised July 3, 1990

We describe a new technique by which single base changes in human genes can be conveniently detected. In this method the DNA fragment of interest is first amplified using the polymerase chain reaction with an oligonucleotide primer biotinylated at its 5'-end. The amplified 5'-biotinylated DNA is immobilized on an avidin matrix and rendered single-stranded. The variable nucleotide in the immobilized DNA is identified by a one-step primer extension reaction directed by a detection step primer, which anneals to the DNA immediately upstream of the site of variation. In this reaction a single labeled nucleoside triphosphate complementary to the nucleotide at the variable site is incorporated. The method is highly sensitive, allowing the use of nucleoside triphosphates labeled with radioisotopes of low specific activity (^3H) as well as nonradioactive markers (digoxigenin). The procedure consists of few and simple operations and is thus applicable to the analysis of large numbers of samples. Here we applied it to the analysis of the three-allelic polymorphism of the human apolipoprotein E gene. We were able to correctly identify all possible combinations of the three apo E alleles. © 1990 Academic Press, Inc.

INTRODUCTION

Changes in only one or a few nucleotides have been found to cause several types of human hereditary diseases (Antonarakis, 1989). The increasing understanding of the exact nature of the genetic defects causing human diseases has created a need for convenient diagnostic methods to detect these defects, at both the prenatal and postnatal stages. Polymerase chain reaction (PCR) amplification (Mullis and Faloona, 1987) has significantly improved both the sensitivity and the specificity of analyzing minute changes in the human genome. A number of techniques have been employed to identify mutations in the amplified DNA. In some cases, nucleotide substitutions in the

amplified fragments may be detected by analysis of restriction site variation (Saiki *et al.*, 1985; Kogan *et al.*, 1987), which is feasible when the nucleotide variation alters a restriction enzyme cleavage site. Dot blot hybridization with sequence-specific oligonucleotide probes is another generally used method (Saiki *et al.*, 1986; Smeets *et al.*, 1988; Farr *et al.*, 1988). Although its use is not limited to the identification of nucleotide changes creating or eliminating restriction sites, this technique does not allow the identification of somatic mutations present in a small fraction of the cell population.

Several strategies have been employed to develop new, more convenient methods for the detection of nucleotide variations in the amplified DNA. In a "reversion" of oligonucleotide hybridization the sequence-specific oligonucleotide probes are immobilized, which allows the simultaneous analysis of one sample with several probes (Saiki *et al.*, 1989). Instead of using allele-specific oligonucleotides to the nucleotide change itself, they have also been used to prime the PCR amplification. In this technique, allele-specific amplification is achieved using a primer with a 3'-mismatch at the position of the variable nucleotide (Wu *et al.*, 1989) or by a competitive priming reaction, in which the mismatch is located within the primer (Gibbs *et al.*, 1989). A ligation-mediated reaction, in which a pair of oligonucleotides annealing at adjacent positions at the site of the nucleotide variation, has been used to analyze PCR-amplified DNA (Landegren *et al.*, 1988; Wu and Wallace, 1989). Attachment of a "GC-clamp" to the 5'-end of the amplified DNA allows the identification of single base substitutions on the basis of altered melting properties of the DNA fragments in denaturing gradient gel electrophoresis (Sheffield *et al.*, 1989). Another approach is to introduce mobility shifting nucleotide analogs into the amplified DNA sample. The nucleotide variation is then identified by observing the mobility of the DNA

fragments using electrophoresis in a denaturing gel (Kornher and Livak, 1989). Chemical cleavage of mismatched duplexes in the amplified DNA has also been used to detect point mutations (Cotton *et al.*, 1988).

When using sequence-specific oligonucleotides as hybridization probes or as PCR primers, the reaction conditions are extremely critical. The gel electrophoretic separation step required in several of the methods described above is inconvenient to carry out for large numbers of samples.

We report a new method for the identification of single base variations in human DNA, in which the above-mentioned limitations are avoided. We applied the method to analyze the genetic polymorphism of the human apolipoprotein E (apo E). Apo E plays an important role in lipoprotein metabolism (Mahley, 1988). It is both an integral component and a mediator of cellular uptake of several lipoproteins. In population studies it has been shown that the polymorphism of the apo E, which is due to single base substitutions at two loci in the apo E gene, can explain as much as 10% of the individual variations in serum cholesterol levels (Davignon *et al.*, 1988). In addition, this polymorphism may also affect the individual risk of atherosclerotic vascular disease (Davignon *et al.*, 1988). Apo E exists as three common isoforms, E2, E3, and E4. These isoforms are encoded by three alleles (ϵ 2, ϵ 3, and ϵ 4) that differ from each other by single base substitutions in the codons for amino acid residues 112 and 158. The frequencies of the ϵ 2, ϵ 3, and ϵ 4 alleles are 10, 75, and 15%, respectively (Mahley, 1988). We reasoned that a convenient DNA technique for apo E genotyping would be very useful in genetic-epidemiological studies of hyperlipoproteinemias and may also have impact for the early detection of individuals with increased risk of cardiovascular disease.

MATERIALS AND METHODS

Clinical Samples

Venous blood samples were obtained from patients attending the Lipid Outpatient Clinic of the University Central Hospital of Helsinki. Apo E phenotyping was accomplished by isoelectric focusing (Ehnholm *et al.*, 1986). Leukocytic DNA was isolated according to Bell *et al.* (1981).

Preparation of the Primers

Four PCR primers (denoted P1-P4) and two detection step primers (D112 and D158) were synthesized on an Applied Biosystems 381A DNA synthesizer (Beaucage and Caruthers, 1982). The primers were designed on the basis of the known nucleotide se-

quence of the apo E gene (Paik *et al.*, 1985). Their sequences and positions on the apo E gene are given in Table 1. A 5'-amino group was added to the primer P2 with the Aminolink II reagent (Applied Biosystems). A biotin residue was attached to the amino group using a water soluble sulfo-NHS-biotin ester (Pierce Chemical Co.) and the biotinylated oligonucleotide was purified by reversed phase HPLC (Bengtström *et al.*, 1990).

Polymerase Chain Reaction

The DNA (100 ng per sample) was amplified with the P1 and P4 primers (final concentrations 1 μ M) in 100 μ l of a solution of 0.2 mM each of dATP, dCTP, dGTP, dTTP, 20 mM Tris-HCl, pH 8.8, 15 mM $(\text{NH}_4)_2\text{SO}_4$, 1.5 mM MgCl_2 , 0.1% Tween 20, 0.01% gelatin, and 2.5 units of *Thermus aquaticus* (Taq) DNA polymerase (United States Biochemical Corp.) in a DNA thermal cycler (Perkin-Elmer/Cetus) for 25 cycles for 1 min at 96°C and 2 min at 65°C. For reamplification with a nested pair of primers, an aliquot (3 μ l of a 1:100 dilution) of the first PCR product amplified with P1 and P4 was transferred to a second PCR. This was carried out under the conditions described above and directed by the biotinylated primer P2 and the primer P3 at 1 or 0.1 μ M concentration.

Affinity-Capture of the Biotinylated Amplified DNA on Avidin-Coated Polystyrene Particles

Ten microliters of a 5% (w/v) suspension of avidin-coated polystyrene particles (0.8 μ m, Baxter Healthcare Corp.) was added to an 80- μ l aliquot of the PCR mixture. The samples were kept at 20°C for 30 min. The particles were collected by centrifugation for 2 min in an Eppendorf centrifuge at 6000g and were washed once by vortexing vigorously for a few seconds with 1 ml of 15 mM NaCl, 1.5 mM Na-citrate (0.1 \times SSC), 0.2% sodium dodecyl sulfate (SDS), and once with 1 ml of 0.1% Tween 20 in 0.15 M NaCl, 20 mM phosphate buffer, pH 7.5 (PBS). The particles were treated twice with 200 μ l of 0.15 M NaOH for 5 min at 20°C. The particles were then washed once with 1 ml of 0.1% Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5, and twice with 1 ml of 0.01% Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5. The particles are easy to handle when the solutions contain detergent and the concentration of NaCl is not above 0.15 M.

The suspension of particles in the last washing solution was divided into two or four aliquots depending on the labeling system to be used. The particles were collected by centrifugation in separate tubes.

The Detection Step Primer Extension Reaction on Microparticles

The particles carrying the DNA fragment were suspended in 10 μ l of 50 mM NaCl, 20 mM MgCl_2 , 40 mM

Tris-HCl, pH 7.5, containing 2 pmol of the D112 or the D158 primer. The primer was allowed to anneal to the DNA template by heating the samples at 65°C for 2 min and cooling them to 20°C during the next 30 min. One microliter of 0.1 M dithiothreitol (DTT) and a mixture of the appropriate deoxynucleoside triphosphates (dNTPs) and dideoxynucleoside triphosphates (ddNTPs) were added to yield 1 μ M concentrations each in a final volume of 15 μ l. For identification of T, [³⁵S]dTTP (600 Ci/mmol, Amersham; 6 pmol diluted with 9 pmol of unlabeled dTTP), ddCTP, and ddGTP were added to one sample. For the identification of C, [³⁵S]dCTP (1000 Ci/mmol, Amersham; 5 pmol diluted with 10 pmol of unlabeled dCTP), ddTTP, and ddGTP were added to another sample. Alternatively, [³H]dTTP (114 Ci/mmol, Amersham) and [³²P]dCTP (3000 Ci/mmol, Amersham), diluted in unlabeled dCTP to a specific activity of 150 Ci/mmol, and ddGTP were added to one sample. Two microliters (3 units) of T7 DNA polymerase (Sequenase, United States Biochemical Corp.) was added to each tube and the reaction was allowed to proceed for 6 min at 42°C.

For comparison, *Taq* DNA polymerase and *Escherichia coli* DNA polymerase I (the Klenow fragment) were also used. The reaction with the *Taq* DNA polymerase was carried out at 55°C in the PCR buffer. A 60-fold excess of ddNTP was added to the reaction with the Klenow DNA polymerase and the buffer contained 50 mM Tris-HCl, pH 8.1, 2 mM dithiothreitol, 5 mM MgCl₂, 40 mM KCl.

After the incorporation step the microparticles were washed twice with 1 ml of 0.1× SSC, 0.2% SDS, and twice with 0.1% Tween 20 in PBS at 20°C. For elution of the reaction products the particles were boiled in 200 μ l of H₂O for 5 min, cooled on ice, and centrifuged for 2 min in an Eppendorf centrifuge. The eluted radioactivity was measured in a liquid scintillation counter (Rackbeta 1219, Pharmacia/Wallac). ³H and ³²P were measured simultaneously by setting the window for ³H at channels 10–90 and the window for ³²P at channels 130–220.

Affinity-Capture of the Amplified DNA in Avidin-Coated Microtitration Wells

Two 15- μ l aliquots of the second PCR mixture (amplified using 0.1 μ M concentrations of primers) were transferred to microtitration wells (Nunc, Maxisorb) that had been coated with streptavidin by passive adsorption. Thirty microliters of 0.1% Tween 20 in 0.15 M NaCl, 0.1 M Tris-HCl, pH 7.5 (TBS), was added to each well. The microtitration strips were incubated for 3 h at 37°C with gentle shaking. The wells were washed three times with 200 μ l of 0.1% Tween 20 in TBS at 20°C. The wells were then treated twice with

100 μ l of 50 mM NaOH for 5 min at 20°C, followed by washing twice with 200 μ l of 0.1× SSC, 0.2% SDS, twice with 0.1% Tween 20 in TBS, once with 0.1% Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5, and finally once with 0.01% Tween 20 in 50 mM NaCl, 40 mM Tris-HCl, pH 7.5.

The Detection Step Primer Extension Reaction in Microtitration Wells

Ten picomoles of the primer D112 or D158 was added to each well in 50 μ l of 0.9 M NaCl, 0.2 M Tris-HCl, pH 7.5. The wells were heated for 2 min at 65°C and allowed to cool to 20°C during 30 min. The mixture was discarded and the wells were washed once with 200 μ l of 0.25 M NaCl, 0.2 M Tris-HCl, pH 7.5, at 20°C. Fifty microliters of a solution consisting of 1 μ M digoxigenin-11-dUTP (Boehringer-Mannheim), 1 μ M ddCTP, 1 μ M ddGTP, 0.2 μ M primer (D112 or D158), 6 mM dithiothreitol, 37.5 mM NaCl, 15 mM MgCl₂, 30 mM Tris-HCl, pH 7.5, and 3 units of T7 DNA polymerase was added. The microtitration strips were incubated for 10 min at 42°C, and the wells were washed twice with 200 μ l of 0.1× SSC, 0.2% SDS and three times with 200 μ l of 0.1% Tween 20 in TBS. Then, 60 μ l of a 1:1000 dilution of an anti-digoxigenin-alkaline phosphatase conjugate (Boehringer-Mannheim) in a solution of 0.1% Tween 20, 1% bovine serum albumin in TBS was added, and the microtitration strips were incubated for 2 h at 37°C with gentle shaking. The wells were washed six times with 0.1% Tween 20 in TBS and once with 1 mM diethanolamine–0.5 mM MgCl₂ buffer, pH 10. Finally 160 μ l of 2 mM *p*-nitrophenyl phosphate in the alkaline buffer was added. After development of color for 20 min at room temperature the absorbance of the formed product was measured at 405 nm in a spectrophotometric reader.

RESULTS AND DISCUSSION

Principle of the Method

In the present technique the nucleotide at the variable site of the target DNA is detected by a one-step primer extension reaction. A specific primer, which anneals to the target DNA immediately upstream of the variable site, is elongated with a single labeled nucleoside triphosphate complementary to the variable nucleotide (Fig. 1).

Prior to the detection step (primer extension reaction) the target DNA is immobilized and rendered single-stranded. This is accomplished by first amplifying the DNA region containing the variable nucleotide using a 5'-biotinylated PCR primer (Syvänen *et al.*, 1988). The synthesized DNA fragment carrying a 5'-biotin residue on one of the strands is then cap-

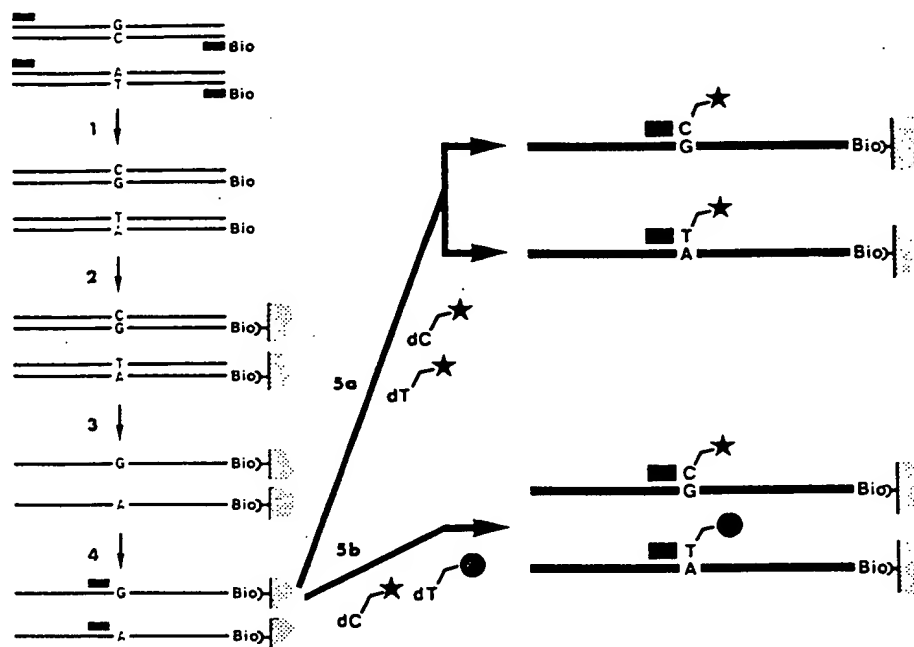


FIG. 1. Principle of the method. 1. Amplification with one biotinylated PCR primer. 2. Affinity-capture. 3. Washing and denaturation. 4. Annealing of the detection step primer. 5. The detection step reaction: a, one label, divided sample; b, two labels, undivided sample. Bio denotes biotin; A symbolizes avidin or streptavidin.

tured on a solid matrix, taking advantage of the interaction between biotin and avidin (Syvänen *et al.*, 1989). Immobilization of the amplified DNA enables efficient removal of the excess of PCR primers and dNTPs, as well as denaturation of the double-stranded amplified DNA fragment. This is a prerequisite for carrying out the detection step primer extension reaction.

If the two dNTPs used to detect the nucleotide variation carry the same label, the sample is divided before the detection step reaction. Alternatively an undivided sample is analyzed with two dNTPs carrying different labels (Fig. 1). The procedure allows the identification of nucleotide substitutions in samples from both homozygous and heterozygous individuals. In a sample from a heterozygote, a signal is obtained from the reactions with both labeled dNTPs. When a sample from a homozygous individual is analyzed, only one signal corresponding to the nucleotide present is obtained.

Here we applied the method to analyze the three-allelic polymorphism of the apo E gene. The polymorphism of apo E is due to single base substitutions in the codons for amino acids 112 and 158. The nucleotides of codons 112 and 158 are either CGC (encoding arginine) or TGC (encoding cysteine) (Fig. 2). The nucleotides at both variable sites were determined from the same amplified fragment. For this the PCR-amplified sample was divided and analyzed using a de-

tection step primer specific for codon 112 and another detection step primer specific for codon 158.

Preparation of Immobilized Single-Stranded DNA

The specificity of the method is affected by the quality and quantity of the biotinylated PCR product. We used two consecutive PCRs with nested sets of primers. The primary primers (P1 and P4) were 328 bp apart on exon 4 of the apo E gene. The nested primers P2 (biotinylated) and P3 amplified a 265-bp fragment over the region coding for amino acids 112 and 158. For comparison, a single PCR with the latter primer pair was carried out directly on the genomic DNA (Fig. 3).

Amino acid position:	
112	158
E2 -- Cys	Cys --
E2 -- TGC	TGC --
E3 -- Cys	Arg --
E3 -- TGC	CGC --
E4 -- Arg	Arg --
E4 -- CGC	CGC --

FIG. 2. The apolipoprotein E isoforms and alleles. The protein isoforms are designated E2, E3, and E4. The corresponding alleles are ϵ_2 , ϵ_3 , and ϵ_4 . The amino acid residues at positions 112 and 158 and the corresponding codons are shown.

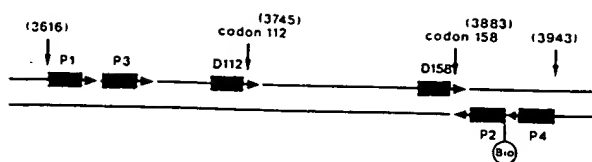


FIG. 3. Location of the polymerase chain reaction and detection step primers on the apolipoprotein E gene. P1-P4 are PCR primers; D112 and D158 are the detection step primers for codons 112 and 158, respectively. The numbers in parentheses refer to nucleotide numbers on the apo E gene according to Paik *et al.* (20). The drawing is to scale. The arrows indicate the direction of the primer extension reaction. Bio denotes biotin.

The quality of the PCR product was assessed by gel electrophoresis of 1/10th of a PCR reaction. Using the nested PCR method a single band was always observed, while after a single PCR a few extra bands were occasionally seen. The amount of PCR product was estimated by the affinity-based hybrid collection method (Syvänen *et al.*, 1988). In the two-step PCR we obtained about 5 pmol of product from the reactions, whereas the amount of amplified DNA varied between 0.1 and 0.8 pmol when only a single PCR was used.

The biotinylated PCR-amplified DNA was captured on avidin-coated polystyrene particles or in microtitration wells coated with streptavidin. The biotin binding capacity of the amount of avidin particles used is sufficient to capture at least 100 pmol of biotin (Syvänen *et al.*, 1988). The alkaline treatment used to remove the nonbiotinylated strand of the amplified DNA does not disrupt the biotin-avidin interaction. The biotin binding capacity of the microtitration wells is lower than that of the microparticles (Harju *et al.*, 1990). The amount of biotinylated primer in the second PCR was therefore reduced to 10 pmol when microtitration wells were used to capture the amplified DNA.

The Detection Step Reaction

The detection step primers D112 and D158 are 20-mer hybridizing to the region immediately upstream of the variable first nucleotide of codons 112 and 158, respectively (Table 1, Fig. 3).

The effect of the amount of primer used in the detection step reaction was tested. Increasing the amount of primer from equal molar concentration to 2-, 5-, and 10-fold molar excess in relation to immobilized target DNA did not have any effect on the signal or the specificity of the reaction. Omitting the separate annealing reaction decreased the efficiency of the reaction to one-fourth of that obtained with the standard technique, i.e., with consecutive primer-anneal-

TABLE 1
Nucleotide Sequence and Position on the Apolipoprotein E Gene of the Oligonucleotide Primers

Primer	Sequence	Position ^a
P1	5'-AAG GAG TTG AAG GCC TAC AAA T	3616-3677
P3	5'-GAA CAA CTG AGC CCG GTG GCG G	3649-3677
D112	5'-GCG CGG ACA TGG AGG ACG TG	3725-3744
D158	5'-ATG CCG ATG ACC TGC AGA AG	3863-3882
P2	5'-TCG CGG GCC CCG GCC TGG TAC A	3914-3893
P4	5'-GGA TGG CGC TGA GGC CGC GCT C	3943-3922

^a The positions are given as nucleotide numbers of the apo E gene according to Paik *et al.* (20).

ing and labeling-termination reactions using T7 DNA polymerase.

The design of the dNTP/ddNTP mixture used in the labeling-termination reaction is an important parameter of this method. Inclusion of ddNTPs into the reaction mix ensures complete termination of the reaction. When C at either of the variable sites in the apo E gene is to be detected, labeled dCTP and unlabeled ddTTP are used; when a T is to be detected, the mixture contains labeled dTTP and unlabeled ddCTP. In addition, ddGTP, which corresponds to the second nucleotide in codons 112 and 158, is included in the reaction mixtures. When dCTP and dTTP carrying different labels are used in the same reaction, only ddGTP is added. The optimal result was obtained with all dNTPs and ddNTPs at 1 μ M concentration.

We compared three DNA polymerases in the labeling-termination reaction. The results presented in Table 2 show that the T7 and *T. aquaticus* DNA poly-

TABLE 2
Comparison of DNA Polymerases in the Detection Step Reaction^a

Sample phenotype ^b	Enzyme ^c	Radioactivity collected (cpm) ^d	% misincorporation
E2/E2 (T/T)	T7	107,000	
E4/E4 (C/C)	T7	1,600	1.5
E2/E2 (T/T)	Taq	137,000	
E4/E4 (C/C)	Taq	1,800	1.3
E2/E2 (T/T)	Klenow	138,000	
E4/E4 (C/C)	Klenow	28,200	21.0

^a The comparison was carried out using 3 units of each enzyme and the detection step primer D158 and [³H]dTTP as label.

^b The first nucleotide of codon 158 is given in parentheses.

^c T7, T7 DNA polymerase; Taq, *Thermus aquaticus* DNA polymerase; Klenow, the Klenow fragment of the *E. coli* DNA polymerase I.

^d The background values from control reactions without DNA (240-380 cpm) have been subtracted.

TABLE 3

Detection of the Variable Nucleotides in Codons 112 and 158 of the Apolipoprotein E gene with ^{35}S -Labeled dTTP and dCTP

No.	Sample Phenotype	Radioactivity incorporated (cpm) ^a				First nucleotide	
		Codon 112		Codon 158		Codon 112	Codon 158
		T-reaction	C-reaction	T-reaction	C-reaction		
1.	E2/E2	105,000	826	42,100	390	T/T	T/T
2.	E3/E2	111,000	680	89,900	16,200	T/T	T/C
3.	E4/E2	38,500	14,100	27,900	5,720	T/C	T/C
4.	E3/E3	86,500	444	655	43,100	T/T	C/C
5.	E4/E3	50,500	8,370	245	15,300	T/C	C/C
6.	E4/E4	1,050	32,100	395	15,000	C/C	C/C

^a The background values from control reactions without DNA have been subtracted (Codon 112: T-reaction, 502 cpm; C-reaction, 214 cpm. Codon 158: T-reaction, 575 cpm; C-reaction, 790 cpm).

merases perform satisfactorily. The *E. coli* DNA polymerase I (the Klenow fragment) yielded significant misincorporation of labeled dNTP. The reason for this is probably the 3' → 5' exonuclease proofreading activity of the Klenow DNA polymerase, which is lacking in the two other enzymes (Tabor and Richardson, 1987; Innis *et al.*, 1988).

Detection of the Polymorphism of the Apolipoprotein E Gene

Using the optimized procedure we analyzed six samples, which by isoelectric focusing were first shown to correspond to the six possible combinations of the three apo E alleles.

In one series, dNTPs labeled with ^{35}S were used. Each sample, amplified by the "nested" PCR method, was collected on avidin-coated microparticles and divided into four aliquots for the detection step reaction. The variable nucleotide (C or T) in the first position of codon 112 was analyzed in two of the aliquots using the primer D112 and the marker molecule [^{35}S]dCTP or [^{35}S]dTTP, respectively. Similar conditions and the primer D158 were used to identify the variable nucleotide in codon 158. The high positive signals and low levels of misincorporated label allowed the unequivocal identification of a C or a T in both variable sites in all six genotypes (Table 3). Heterozygous subjects were identified by a positive signal from both the C- and the T-reaction at the same site (samples no 2, 3, and 5). The C-reactions constantly yielded lower signals than the T-reactions despite the fact that the input of radioactivity was higher in the C-reactions. The reason for this is presumably that the enzyme preferentially incorporates unlabeled dCTP over thio- ^{35}S]dCTP.

By combining PCR amplification with the simple solid-phase detection step reaction a highly sensitive

and specific technique for identification of single nucleotide variations in the human genes was thus constructed. From the signals obtained in the experiment presented in Table 3, in which each amplified sample was divided into four aliquots, it can be calculated that mutations present in less than a 1% minority of a cell population can be identified. A sensitivity of this range is required when specific minority point mutations, such as those of the *ras* oncogene family (Bos *et al.*, 1987; Farr *et al.*, 1988), are to be detected in cell or tissue samples containing mutant alleles in the presence of an excess of the normal gene.

The apo E gene is a particularly favorable target for analysis by this method for two reasons. First, the close proximity of the variable DNA sites allows their co-amplification within a single DNA fragment. Second, only three allelic forms are prevalent in the population, which implies that all three alleles can be unequivocally identified. When more than one variant DNA site is to be detected and more than three allelic combinations exist, assignment of each variable nucleotide to the correct allele is beyond the ability of the present technique. This problem could be circumvented by allele-specific amplification of the DNA fragment (Wu *et al.*, 1989; Gibbs *et al.*, 1989) with a biotinylated PCR primer prior to the detection step reaction.

Here we applied the technique for the detection of single base substitutions in the human genome. Deletion and insertion mutations in the target DNA can be detected equally well. The only restriction is that the first nucleotide of the deletion/insertion must not be identical to the first nucleotide after the deletion/insertion.

Double-Labeling Systems

A modification of the technique described above and involving the use of a double-labeling system was

TABLE 4

Detection of the Variable Nucleotides in Codons 112 and 158 of the Apolipoprotein E Gene with ^3H -labeled dTTP and ^{32}P -Labeled dCTP

Sample		Radioactivity incorporated (cpm) ^a				First nucleotide	
		Codon 112		Codon 158			
		T-reaction	C-reaction	T-reaction	C-reaction		
No.	Phenotype	T-reaction	C-reaction	T-reaction	C-reaction	Codon 112	Codon 158
1.	E2/E2	29,300	286	33,200	722	T/T	T/T
2.	E3/E2	18,400	309	44,600	28,500	T/T	T/C
3.	E4/E2	10,600	3,840	16,100	7,300	T/C	T/C
4.	E3/E3	16,400	72	216	26,400	T/T	C/C
5.	E4/E3	2,640	7,470	455	22,500	T/C	C/C
6.	E4/E4	172	10,000	158	16,500	C/C	C/C

^a The cpm values from control reactions without DNA have been subtracted (Codon 112: T-reaction, 215 cpm; C-reaction, 379 cpm. Codon 158: T-reaction, 80 cpm; C-reaction, 65 cpm).

also developed. In the modified technique, the immobilized, amplified sample was divided into two aliquots: one for analysis with the D112 and one for analysis with the D158 primer. ^3H -labeled dTTP and ^{32}P -labeled dCTP, diluted in unlabeled dCTP to yield a specific activity similar to that of the [^3H]dTTP, were used as markers in the same reaction. After completion of the detection step reaction, the radioactivity resulting from the incorporated ^3H - and ^{32}P -isotope was measured simultaneously in the liquid scintillation counter. Again, the variable nucleotides at both sites were easily identified in all six samples (Table 4).

The fact that the variable sites of the apo E gene are located only 138 bases apart enables detection of the nucleotide variations at both loci from the same amplified fragment. To identify nucleotide variations located too distant from each other to be amplified with one primer pair, PCR can be carried out using multiple primer pairs in the same reaction. Such a "multiplex" PCR has been used to simultaneously amplify six distant loci in the muscular dystrophy gene (Chamberlain *et al.*, 1989). The sensitivity of our method allows division of one sample into a large number of aliquots for analysis of several sites with different detection step primers. In such cases, the use of a double-labeling system is especially advantageous by halving the number of aliquots to be analyzed.

The high sensitivity of detection allows the use of ^3H -labeled dNTPs, which have low specific activities and the advantages of weak β -emission and a long half-life (13 years).

Stable Labels

The applicability of nonradioactive detection of the different apo E genotypes was also demonstrated (Table 5). In this case, the amplified DNA was collected

in microtitration wells coated with streptavidin. Digoxigenin-11-dUTP was used to identify T in codons 112 and 158. The incorporated digoxigenin-11-dUTP was detected enzymatically by an anti-digoxigenin-alkaline phosphatase conjugate. The colored end-product of the enzymatic reaction was measured in a spectrophotometric reader. The absorbance signals obtained from the samples with a T as the first nucleotide in codon 112 (samples 1–5) and codon 158 (samples 1–3) were clearly distinguishable from the signals generated by the samples with only a C at the variable site (Table 5). High positive signals were obtained despite the fact that only a small portion (15%) of the PCR sample was analyzed per microtitration well. Other dNTPs modified with haptens such as fluorescein isothiocyanate or dinitrophenyl groups (Kumar *et al.*, 1988; Vincent *et al.*, 1982) may analogously be used. To avoid the indirect detection of a hapten with a labeled antibody, directly detectable nonradioactive labels would greatly simplify the pro-

TABLE 5

Nonradioactive Identification of T in Codons 112 and 158 of the Apolipoprotein E Gene

Sample		Absorbance at 405 nm ^a		Identification of T	
No.	Phenotype	Codon 112	Codon 158	Codon 112	Codon 158
1.	E2/E2	1.29	1.31	+	+
2.	E3/E2	1.31	0.79	+	+
3.	E4/E2	0.32	0.28	+	+
4.	E3/E3	0.41	0.06	+	–
5.	E4/E3	0.65	0.07	+	–
6.	E4/E4	0.06	0.09	–	–

^a The absorbance readings from control reactions without DNA have been subtracted (codon 112, 0.09; codon 158, 0.11).

cedure. Any label that can be attached to a nucleoside triphosphate in such a way that the nucleotide is incorporated in a sequence-specific fashion can be used. In preliminary experiments we have shown that ddNTPs modified with derivatives of fluorescein isothiocyanate (Prober *et al.*, 1987) are applicable to the present method (Syvänen *et al.*, unpublished). The sensitivity of detection in a standard fluorometer (Hitachi, F-1000) is sufficient. In comparison with dNTPs, labeled ddNTPs provide an additional advantage in that the labeling and termination steps are combined into the same reaction. Thus, double-labeling systems can be applied even in cases where the nucleotide following the variable site is identical to one of the nucleotides at the variable site.

CONCLUSIONS

Compared to other methods used to detect nucleotide changes in the human genome the method described here has several major advantages. It comprises few and simple steps which all are carried out in test tubes or microtitration wells. Because the target DNA is immobilized before analysis, electrophoretic separation is avoided. The results are obtained as numeric values, enabling objective interpretation of the data. Thus, the method consists of elements suitable for automatization with already existing laboratory robots (Wilson *et al.*, 1988). Furthermore, the high sensitivity of detection allows the identification of mutations present in only a small fraction of the analyzed cells. Nonradioactive detection methods are applicable, which renders the technique suitable to hospital and other clinical laboratories. These facts make the primer-guided nucleotide incorporation assay a promising tool for the diagnosis of inherited diseases and for the detection of somatic mutations.

ACKNOWLEDGMENTS

We thank Anne Järvinen for excellent technical assistance, Marina Bengtström and Maritta Putkiranta for providing the biotinylated oligonucleotide, Pentti Somerharju for help with fluorescence measurements, and Jukka Tenhunen for drawing the figures. This work was supported by grants from the Finnish Academy of Sciences, The University of Helsinki, and the Sigrid Jusélius Foundation.

REFERENCES

- ANTONARAKIS, S. E. (1989). Diagnosis of genetic disorders at the DNA level. *N. Engl. J. Med.* 320: 153-163.
- BEAUCAGE, S. L., AND CARUTHERS, M. H. (1982). Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* 22: 1859-1862.
- BELL, G. I., KARAM, J. H., AND RUTTER, W. J. (1981). Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Proc. Natl. Acad. Sci. USA* 78: 5759-5763.
- BENGSTSTRÖM, M., JUNGELL-NORTAMO, A., AND SYVÄNEN, A.-C. (1990). Biotinylation of oligonucleotides using a water soluble biotin ester. *Nucleosides Nucleotides* 9: 123-127.
- BOS, J. L., FEARON, E. R., HAMILTON, S. R., VERLAAN-DE VRIES, M., VAN BOOM, J. H., VAN DER EB, A. J., AND VOGELSTEIN, B. (1987). Prevalence of ras gene mutations in human colorectal cancer. *Nature (London)* 327: 293-297.
- CHAMBERLAIN, J. S., GIBBS, R. A., RANIER, J. E., NGUYEN, P. N., AND CASKEY, C. T. (1989). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res.* 16: 11,141-11,156.
- COTTON, R. G. H., RODRIGUES, N. R., AND CAMPBELL, R. D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA* 85: 4397-4401.
- DAVIGNON, J., GREGG, R. E., AND SING, C. F. (1988). Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 8: 1-21.
- EHNHOLM, C., LUKKA, M., KUUSI, T., NIKKILÄ, E., AND UTERMANN, G. (1986). Apolipoprotein E polymorphism in the Finnish population: Gene frequencies and relation to lipoprotein concentrations. *J. Lipid Res.* 27: 227-235.
- FARR, C. J., SAIKI, R. K., ERLICH, H. A., MCCORMICK, F., AND MARSHALL, C. J. (1988). Analysis of RAS gene mutations in acute myeloid leukemia by polymerase chain reaction and oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 85: 1629-1633.
- GIBBS, R. A., NGYEN, P.-N., AND CASKEY, T. C. (1989). Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Res.* 17: 2437-2448.
- HARJU, L., JÄNNE, P., KALLIO, A., LAUKKANEN, M.-L., LAUTENSCHLAGER, I., MATTINEN, S., RANKI, A.-M., RANKI, M., SOARES, V. R. X., SÖDERLUND, H., AND SYVÄNEN, A.-C. (1990). Affinity-based collection of amplified viral DNA: Application to the detection of human immunodeficiency virus type 1, human cytomegalovirus and human papillomavirus type 16. *Mol. Cell. Probes* 4, 223-235.
- INNIS, M. A., MYAMBO, K. B., GELFAND, D. H., AND BROW, M. A. D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. USA* 85: 9436-9440.
- KOGAN, S. C., DOHERTY, M., AND GITSHIER, J. (1987). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317: 985-990.
- KORNER, J. S., AND LIVAK, K. J. (1989). Mutation detection using nucleotide analogs that alter electrophoretic mobility. *Nucleic Acids Res.* 17: 7779-7783.
- KUMAR, A., TCHEN, P., ROULLET, F., AND COHEN, J. (1988). Nonradioactive labelling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase. *Anal. Biochem.* 169: 376-382.
- LANDEGREN, U., KAISER, R., SANDERS, J., AND HOOD, L. (1988). A ligase-mediated gene detection technique. *Science* 241: 1077-1080.
- MAHLEY, R. W. (1988). Apolipoprotein E: Cholesterol transport protein with expanding role in cell biology. *Science* 240: 622-630.
- MULLIS, K. B., AND FALOONA, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *In*

- "Methods in Enzymology" (R. Wu, L. Grossman, and K. Moldave, Eds.), Vol. 155, pp. 335-350. Academic Press, New York.
20. PAIK, Y.-K., CHANG, D. J., REARDON, C. A., DAVIES, G. E., MAHLEY, R. W., AND TAYLOR, J. M. (1985). Nucleotide sequence and structure of the human apolipoprotein E gene. *Proc. Natl. Acad. Sci. USA* 82: 3445-3449.
 21. PROBER, J. M., TRAINOR, G. L., DAM, R. J., HOBBS, F. W., ROBERTSON, C. W., ZAGURSKY, R. J., COCUZZA, A. J., JENSEN, M. A., AND BAUMEISTER, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238: 336-341.
 22. SAIKI, R. K., SCHARF, S., FALOONA, F., MULLIS, K. B., HORN, G. T., ERLICH, H. A., AND ARNHEIM, N. (1985). Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.
 23. SAIKI, R. K., BUGAWAN, T. L., HORN, G. T., MULLIS, K. B., AND ERLICH, H. A. (1986). Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with allele-specific oligonucleotide probes. *Nature (London)* 324: 163-166.
 24. SAIKI, R. K., WALSH, P. S., LEVENSON, C. H., AND ERLICH, H. A. (1989). Genetic analysis of amplified DNA with immobilized sequence specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86: 6230-6234.
 25. SHEFFIELD, V. C., COX, D. R., LERMAN, L. S., AND MYERS, R. M. (1989). Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA* 86: 232-236.
 26. SMEETS, H. J. M., PODDIGHE, J., STUYT, P. M. J., STALENHOEF, A. F. H., ROPERS, H. H., AND WIERINGA, B. (1988). Identification of apolipoprotein E polymorphism by using synthetic oligonucleotides. *J. Lipids Res.* 29: 1231-1237.
 27. SYVÄNEN, A.-C., BENGSTSTRÖM, M., TENHUNEN, J., AND SÖDERLUND, H. (1988). Quantification of polymerase chain reaction products by affinity-based hybrid collection. *Nucleic Acids Res.* 16: 11,327-11,338.
 28. SYVÄNEN, A.-C., AALTO-SETÄLÄ, K., KONTULA, K., AND SÖDERLUND, H. (1989). Direct sequencing of affinity-captured amplified human DNA: Application to the detection of apolipoprotein E polymorphism. *FEBS Lett.* 258: 71-74.
 29. TABOR, S., AND RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* 84: 4767-4771.
 30. VINCENT, C., TCHEN, P., COHEN-SOLAL, M., AND KOURILSKY, P. (1982). Synthesis of 8-(2,4 dinitrophenyl 2,6 aminohexyl) amino-adenosine 5'-triphosphate: Biological properties and potential uses. *Nucleic Acids Res.* 10: 6787-6796.
 31. WILSON, R. K., YUEN, A. S., CLARK, S. M., SPENCE, C., ARAKELIAN, P., AND HOOD, L. E. (1988). Automation of dideoxynucleotide DNA sequencing reactions using a robotic workstation. *BioTechniques* 6: 776-787.
 32. WU, D. Y., AND WALLACE, R. B. (1989). The ligation amplification reaction (LAR)—Amplification of specific DNA sequences using sequential rounds of template-dependent ligation. *Genomics* 4: 560-569.
 33. WU, D. Y., UGOZZOLI, L., PAL, B. K., AND WALLACE, R. B. (1989). Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86: 2757-2760.

Overlapping Genomic Sequences: A Treasure Trove of Single-Nucleotide Polymorphisms

Patricia Taillon-Miller,¹ Zhijie Gu,¹ Qun Li,¹ LaDeana Hillier,²
and Pui-Yan Kwok^{1,3}

¹Division of Dermatology and ²Genome Sequencing Center, Washington University School of Medicine,
St. Louis, Missouri 63110 USA

An efficient strategy to develop a dense set of single-nucleotide polymorphism (SNP) markers is to take advantage of the human genome sequencing effort currently under way. Our approach is based on the fact that bacterial artificial chromosomes (BACs) and P1-based artificial chromosomes (PACs) used in long-range sequencing projects come from diploid libraries. If the overlapping clones sequenced are from different lineages, one is comparing the sequences from 2 homologous chromosomes in the overlapping region. We have analyzed in detail every SNP identified while sequencing three sets of overlapping clones found on chromosome 5p15.2, 7q21-7q22, and 13q12-13q13. In the 200.6 kb of DNA sequence analyzed in these overlaps, 153 SNPs were identified. Computer analysis for repetitive elements and suitability for STS development yielded 44 STSs containing 68 SNPs for further study. All 68 SNPs were confirmed to be present in at least one of the three Caucasian, African-American, Hispanic) populations studied. Furthermore, 42 of the SNPs tested (62%) were informative in at least one population, 32 (47%) were informative in two or more populations, and 23 (34%) were informative in all three populations. These results clearly indicate that developing SNP markers from overlapping genomic sequence is highly efficient and cost effective, requiring only the two simple steps of developing STSs around the known SNPs and characterizing them in the appropriate populations.

The sequence data described in this paper have been submitted to the GenBank data library under accession nos. AC003015 (for GSII3423), AC002380 (GS330J10), AC000066 (RG293FII), AC003086 (RG104F04), C002525 (257C22A), and U73331 (96A18A).]

There is increasing agreement that association studies using a set of single-nucleotide polymorphism (SNP) markers across the genome with markers evenly distributed at ~100-kb intervals would provide the necessary power to detect small genetic effects for a given complex disease trait (Collins et al. 1997; Kruglyak 1997). To develop 30,000 or more P markers is a priority of the consortium of National Institutes led by the National Human Genome Research Institute (Marshall 1997). Although frequency of SNPs is approximately 1 in 1000 bp between any two chromosomes (Cooper et al. 1985; Kwok et al. 1996), there are currently no efficient ways to find and map them from scratch.

In general, development of SNP markers requires five different steps: obtain DNA sequence, develop STSs from the DNA sequence, screen STSs for SNPs, characterize SNPs, and map SNPs to specific chromosomal locations. To date, much effort has

been devoted to devising more efficient ways to screen STSs for SNPs and to characterize them. Being largely ignored is the fact that the most costly aspects of developing SNP markers are the obtaining of DNA sequence for STS development initially and mapping the SNPs at the end of the process. We and others have developed various strategies to improve the efficiency of this process by utilizing existing resources to our advantage. For example, we have screened mapped STSs for SNPs, thereby reducing the development of SNPs to two steps (screening for and characterizing SNPs), abrogating the need for genomic DNA sequencing, STS development, or mapping (Kwok et al. 1996). Because the mapped STSs were developed for YAC library screening in physical mapping, they amplified short DNA fragments and screening several of them was required before a SNP could be found. Moreover, as the physical mapping effort is winding down over the next 2 years, this resource will not be available for further SNP marker development.

Fortunately, the human genome sequencing effort that is currently under way provides a better

Corresponding author.

P. Kwok@im.wustl.edu; FAX (314) 362-8159.

way to develop a dense set of SNP markers in the genome. Like screening mapped STSs for SNPs, our approach bypasses the need for DNA sequence acquisition and mapping for SNPs. In addition, it eliminates the polymorphism screening step altogether, leaving only the development of STSs around the SNPs found during the course of genome sequencing and their characterization. The strategy is based on the fact that bacterial artificial chromosomes (BACs) and P1-based artificial chromosomes (PACs), the substrates of choice for long-range sequencing, come from diploid libraries. With an average insert size of ~120 kb, one can expect a significant overlap between clones selected for sequencing at ~100 kb intervals. If the clones are from different libraries (presumably from different individuals), one is comparing the sequences from two lineages in the overlapping region. If the clones are from the same library, there is still a 50% chance that the overlapping clones are derived from different lineages (paternal or maternal). This probability could increase to close to 100% if libraries made from mixtures of individuals are used. Although one is sampling just two copies of the same region for polymorphisms in this approach, the chance of identifying a polymorphism in the region is the same as the heterozygosity of the polymorphism in the population. Therefore, for the more informative markers (defined as having heterozygosity of >30%) there is a >30% chance of identifying them when

sequences of overlapping clones are examined. Even with a minimal 10% overlap when clone libraries with 10-fold genomic redundancy are used to provide sequencing clones, a 6-kb overlap at each end of the clone insert will result. Given the general observation that one informative marker is found in 1.5–2.0-kb in the human genome (Kwok et al. 1996), at least one such polymorphism should be found in each 4.5–6-kb overlap. In practice, the overlaps are much larger than 6 kb because of the stringent requirements of physical mapping by fingerprinting in selecting clones for large-scale genome sequencing (Marra et al. 1997), and as described here, one can almost certainly find informative SNPs in them.

To test the validity of this approach, we have analyzed in detail every SNP identified while sequencing three sets of overlapping clones found on chromosome 5p15.2, 7q21–7q22, and 13q12–13q13. We report here that this approach of SNP marker development is highly efficient and cost effective, requiring only the two simple steps of developing STSs around the known SNPs and characterizing them in the appropriate populations.

RESULTS

In the course of sequencing overlapping BAC and PAC clones from human genomic libraries at the Genome Sequencing Center (GSC) at Washington

Table 1. Results of Analyzing 68 SNPs Found in 3 Overlapping Clones

Chromosome location	Length of overlap (bp)	Polymorphisms found	SNPs ^a	SNPs analyzed ^b	Informative SNPs ^c					
					African-American	Caucasian	Hispanic	populations		
								3	>2	>1
5p15.2	81,830	20	18	10 8 STSs 3892 bp	6	8	6	6	6	8
7q21–7q22	59,048	97	83	20 13 STSs 6360 bp	60% 14	80% 15	60% 17	60% 12	60% 16	80% 18
13q12–q13	59,739	66	52	38 23 STSs 7720 bp	70% 11	75% 8	85% 11	60% 5	80% 10	90% 16
Totals	200,617	183	153	68 44 STSs 18172 bp	29% 31	21% 31	29% 34	13% 23	26% 32	42% 42
					46% 46%	46% 46%	50% 50%	34% 34%	47% 47%	62% 62%

^aSNPs found among all of the polymorphisms identified in the overlap.

^bSNPs analyzed by determining their frequencies in the population pools, after discarding from consideration polymorphisms found in repetitive regions. The number of STSs and base pairs scanned in the process of analyzing the SNPs are also listed.

^cNumber of SNPs (and the proportion among the SNPs analyzed) with allele frequencies of >20% for the minor allele are categorized according to the individual populations studied and the number of populations in which these informative SNPs are found.

University in St. Louis, MO, 183 polymorphisms were found in 200.6 kb in three overlapping regions, for an average of 1 polymorphism every 1.1 kb (see Table 1). In the 81.8-kb overlap on chromosome 5, 20 polymorphisms were identified. Of these, 17 polymorphisms were single-base substitution polymorphisms (85%), 1 was an unique insertion/deletion polymorphism (5%), and 2 were short tandem repeat polymorphisms (STRP) (10%). In the 59.0-kb overlap on chromosome 7, 97 polymorphisms were found, with 83 being substitution polymorphisms (86%), 11 insertion/deletion polymorphisms in a run of a single base such as a poly(A) (11%) and 3 STRPs (3%). In the 59.7-kb overlap on chromosome 13, 66 polymorphisms were found, with 49 substitution polymorphisms (74%), 3 unique insertion/deletions (5%), 12 insertion/deletions in a run of a single base (18%), and 2 STRPs (3%). Overall, there were 153 SNPs (substitution and unique insertion/deletion polymorphisms) at a frequency of 1 per 1.3 kb. In contrast, there were only 7 STRPs (1 per 28.7 kb).

The 153 SNPs were evaluated further for their usefulness as genetic markers. Those found in common repeat regions masked by the GSC during the sequence annotation such as Alu and L1 were discarded. In all, 55 SNPs were eliminated by computer analysis 5/19 (26%) in the chromosome 5 overlap, 43/83 (52%) in the chromosome 7 overlap, and 7/52 (13%) in the chromosome 13 overlap. The oligonucleotide selection program (osp) (Hillier and Green 1991) was used to design primers to amplify each of the 98 remaining SNPs. Thirty SNPs were in regions of DNA in which no suitable amplimers could be found, including SNPs in repeat regions other than Alus and L1s and a small number of PCR failures. In all, 44 STSs were developed to amplify the remaining 68 SNPs (spanning 18,172 bp of DNA sequence). Among them, 16 STSs contained 2 SNPs and 8 STSs contained 3 or more SNPs.

The 68 SNPs were confirmed by sequence analysis with the homozygous CHM1 DNA from a homozygous complete hydatidiform mole and pooled DNA samples from 30 individuals each from the Caucasian, Hispanic, and African-American populations (Kwok et al. 1994; Taillon-Miller et al. 1997). The pooled DNA sequencing approach for allele frequency estimation is highly reproducible and has been found to give estimates of within 5% of the true allele frequency as found by genotyping all the individuals in the population pool (Kwok et al. 1994). Allele frequency estimates of the 68 SNPs revealed that the minor allele in 23 SNPs (34% of the analyzed SNPs) had a frequency >20% in all three

populations (>32% heterozygosity, assuming Hardy-Weinberg equilibrium), 32 (47%) had a frequency >20% in at least two populations, and 42 (62%) had a frequency >20% in at least one population.

In addition, 18 new SNPs were discovered during the course of analyzing the 44 STSs in 18.2 kb of DNA sequence contained in the STSs (found at a rate of 1 per 1.0 kb). Among these, 9 (50%) were found to be informative in one or more populations and 3 (17%, 1 per 2.0 kb) were informative in all three populations.

The 26 SNPs with frequencies >20% in all 3 populations tested are presented in Table 2. Information about the remaining SNPs with their estimated allele frequencies in each of the three populations can be found on our public database (currently under construction) accessible through the internet (<http://www.ibc.wustl.edu/SNP>).

DISCUSSION

The overall results of this study confirmed that the chance of finding informative SNPs by use of overlapping regions of clones sequenced as part of the human genome sequencing project was in line with our expectations. We had expected to find one informative SNP per 4.5–6.0 kb, and we found one informative SNP per 4.8 kb (3.9 kb if those markers discovered during the sequence analysis of STSs for characterization of the SNPs were included). Although it is safe to assume that the ethnic origins of the donors of the BAC and PAC libraries are Caucasian, many of the SNPs found were also polymorphic in the African-American and Hispanic populations. These results point to the fact that the more informative SNPs are more ancient and are therefore informative in most populations (Kimura 1983).

On closer examination, however, it is clear that there is a large variation in the frequency of finding SNPs among the different overlaps. For example, whereas SNPs are found at a rate of 1 per 4.5 kb in the 5q15.2 overlap, the rate is 1 per 0.7 kb in the 7q21–7q22 overlap and 1 per 1.1 kb in the 13q12–13q13. After computer screening, the rate of analyzable SNPs ranges from 1 in 8.1 kb for 5q12.2, to 1 in 3.0 kb for 7q21–7q22, to 1 in 1.6 kb for 13q12–13q13.

Furthermore, the chromosome 7q21–7q22 overlap was unique in that it had an extremely large number of repeat elements within which the bulk of the SNPs were found, leaving only 20 of the 83 SNPs (24%) suitable for sequence analysis. Among the analyzable SNPs, >80% were informative in at least

Table 2. SNPs with Frequencies >0.20 in Three Populations

STS Name	Chromosome Location	STS Size (bp)	STS Primer 1 Sequence	STS Primer 2 Sequence	Sequence Context of Polymorphism	Allele 1	Allele 2	African-American Frequency Allele 1	Hispanic Frequency Allele 1	Caucasian Frequency Allele 1
5p0002-1	5p15.2	819	CTTCCCTAAAC AAAACCTAC	ACAAGTGAAAGG AAATGAGTC	GCACCTAGCTTTGTATTAGTTCAGM AGTCTCCAGAAAGAAATCAA	TCAG	*	0.41	0.48	0.31
5p0003-2	5p15.2	719	GTATTGATAGA CTTGCTTCC	AGCCTACACATT TTCTTCTG	AACATTATTACTCAACATATTCCT GGAACCTGATTTCATAAT	C	T	0.55	0.62	0.38
5p0003-4	5p15.2	719	GTATTGATAGA CTTGCTTCC	AGCCTACACATT TTCTTCTG	CCTAAACCTTTCTAGCTTTGATTTG TAACAAAAAACTCTAC	A	T	0.88	0.55	0.65
5p0004-1	5p15.2	307	TTCTTCGTGC TTGACAAAG	TGCATCTTAATC CACTCAAC	AACATGTGCACAAAGAACTCTA/GTGTG GTCAGCGAAATCATGAGT	G	A	0.79	0.63	0.56
5p0004-2	5p15.2	307	TTCTTCGTGC TTGACAAAG	TGCATCTTAATC CACTCAAC	AGAAAGAGAAAGATAGTATTCTC/TCT GTTGTCTTACCAGGAACA	C	T	0.61	0.51	0.47
5p0008-1	5p15.2	137	OTTACTTGATC TGCATGG	AGTAGGTAAGG AAGAGAG	TGCCCTTCTTGCAAGTACTA/G/GGA CACTCCAATCCCATTTCT	C	G	0.63	0.62	0.68
5p0009-1	5p15.2	756	GGAGACAATA CAGATGAG	TCACATCCTGA GAGATTG	TTTTTAATTAAGTGAATTC/TCTACT GACCAACCCACATAACAA	T	C	0.46	0.62	0.34
7q0001-1	7q21-22	627	TGAACGATTC GCAGATTG	TACTTCATGCAC CACCTTC	GTTAAACTATTGTTCTGGTAC/GA AGGAGGGAATGTGAGAC	A	C	0.43	0.48	0.34
7q0002-1	7q21-22	170	AAACATGATT ACAGTGGG	TCTCTCTTTTC TTCTCTC	AATGGAGCAGAAACTGGAAATTC/GT GGGTTTAAAGAAAGGCTG	T	C	0.26	0.22	0.37
7q0003-1	7q21-22	398	ACTTATACCT CCTTAAC	CCAGTGTCAA AGGTATC	AACAGAGGATCTGAGGTGA/GTAT AATTAAAGGAAACAGAT	G	T	0.72	0.37	0.59
7q0004-1	7q21-22	373	TTTCAAGAAAG ATCATGGG	GAAGAGAAAT GGCTTGAG	ATCCTGCTTGAGCAATATTCTC/GAG TCTTGATGAATGTAGTTC	C	G	0.72	0.71	0.70
7q0004-2	7q21-22	373	TTTCAAGAAAG ATCATGGG	GAAGAGAAAT GGCTTGAG	CAGAGAATTTGCATTTAGCTA/CCTAT GACTTGCTGAGCAACA	A	C	0.80	0.75	0.57
7q0006-1	7q21-22	918	TATAATCCCAG CTACTCAG	CTACGTTTTCAT AGACAATTC	TCCAGCCTGGGAGACAGAGG/GA AACTCTGCTCAAAAAGAA	A	G	0.45	0.28	0.34
7q0007-1	7q21-22	539	TAGCATGAAC TCAGAGG	CGATGAATTAAC AGAGCC	CTCTAAGCCCAAGTAACACTA/C/GG CTTTGGGAGAGAAACAGC	C	G	0.43	0.44	0.36

Table 2. (Continued)

STS Name	Chromosome Location	STS Size (bp)	STS Primer 1 Sequence	STS Primer 2 Sequence	Sequence Context of Polymorphism	Allele 1	Allele 2	African-American Frequency Allele 1	Hispanic Frequency Allele 1	Caucasian Frequency Allele 1
7q0007-2	7q21-22	539	TAGCATGAACCT TCAGAAAG	CGATGAATTAAC AGAGCC	TGTTCTGCTCAAGTATTCTA/C/GJATT ATTATGTTAATCTGTTT	C	G	0.54	0.35	0.43
7q0008-1	7q21-22	274	ATAATGGGCAT TATTTGCTG	ACTTTCTTAGCT GTGGAAC	ACAAAAAAGGATGGTCT/C/TJCA TGCAAGCTGTATATTGAT	C	T	0.80	0.75	0.75
7q0011-1	7q21-22	161	ATTCAAAGGGT GAAATAAGG	TAGAGAGAAAG AGAGTGAG	TAAGGTAGGCAATTTTAATA/A/GJGC CTCAGAATTTTAATTGA	A	G	0.46	0.29	0.38
7q0011-2	7q21-22	161	ATTCAAAGGGT GAAATAAGG	TAGAGAGAAAG AGAGTGAG	TTAAATTTTTTAATGTTCT/A/GJAGCT TTCTCTTCTTTCTTC	A	G	0.56	0.3	0.38
7q0012-1	7q21-22	611	GATCCTTTGAA TTATTCGTG	ATCATCTTAGCA AAGTGCC	TAAATTAATAATTTGTCTCA/C/AJACG CTTCAGTGAGCCAAGA	C	A	0.67	0.33	0.46
13q0003-2	13q12-q13	170	TGGAATTGAG ATACAGTGG	TGCATACACAC AATTAGGTC	CCCTGAAACTCGATGCAATG/C/GJTT GACITATGAAGAGACCT	C	G	0.77	0.48	0.58
13q0004-2	13q12-q13	241	CAAGGTCATG TATCTAAGG	TCTAAGAAAGC AGTAGCAC	CCCTCGTAAGTTCCTTAGGC/A/GJTT TTTATTCCTTTGTTCAAC	A	G	0.76	0.53	0.54
13q0006-1	13q12-q13	352	CAGCACATTTC AATTCAGC	GAGTCAGAAAC CCTTATATG	CAATTCAGCTGTGTTTCA/C/TJGTGCT CAGTAGCACATGTAC	C	T	0.23	0.71	0.23
13q0013-3	13q12-q13	858	GAAACTGAAG TACAAAGGG	ACCTGGCTAAG AAATGAATC	TAGGATGAAATACAAATAAAC/C/TJGA GCATAAACTTTTTCAC	C	T	0.57	0.72	0.46
13q0020-3	13q12-q13	783	CAGGAACAAAT ACTAGCAAC	GCAGAAACAGA ATATCCTTG	TCCTCCTCCCGTTTCTGTTTTC/TJ GTGCTCACACAGACATAT	TC	*	0.41	0.49	0.48
13q0021-2	13q12-q13	614	GAATCACAA GGTTTGAGG	CGAGAACTCTA ATTCAAGG	GCAGCACTTACTTTCAGCTGT/CJCC AGCTCTAAGCCAGAGTTC	T	C	0.42	0.68	0.31
13q0022-1	13q12-q13	261	ATATTCCTTCC TCTGCCTG	TCACTATGTATG GTGAGTC	TTCCCAAGAGATTAAATAT/T/CJGCT TTGCAGCTTA TTATCTC	T	C	0.53	0.69	0.45

*GenBank accession nos. are AC003015 (for GS113H23) and AC002380 (for GS33010) on 5p15.2; AC000066 (for RG293F11) and AC003086 (for RG104F04) on chromosome 7q21-7q22; AC002525 (for 257C22A) and U73331 (for 96A18A) on chromosome 13q12-13q13.

one population in both the chromosome 5 and 7 overlaps. In contrast, 38 of the 58 SNPs (73%) in the chromosome 13q12-13q13 overlap were analyzable but only 16 of 38 SNPs (42%) were informative.

Despite these regional differences, however, the overall results show that even in the worst case scenario, one only has to analyze two to three SNPs to find an informative SNP marker. Given that almost all overlaps produced by large-scale genome sequencing projects are >20 kb, there will be more than enough analyzable SNPs to choose from.

Our approach has many advantages not found in other current methods. First, all long-range sequencing groups produce high quality sequence data, and because every base is sequenced at least twice from each clone (so-called double-stranding), the error rate is therefore much lower than the polymorphism rate. Second, because the polymorphism data are generated by examining existing data, the amount of sequencing required is minimal and the cost of the project is shifted from identification of polymorphisms toward estimates of the usefulness of the polymorphism by population sequencing. Third, because they are derived from long-range sequence data, the markers are precisely mapped, not just assigned to an interval of a clone-based contig by STS content mapping. Fourth, the physical distance between markers is known precisely. Fifth, because they are detected when only two chromosomes are examined, each SNP identified in this way has a higher chance of being informative. Sixth, this approach scales easily because the basic methods used are simple and robust, making it possible to keep up with the expanding sequencing efforts around the world. Consequently, the genetic map could be completed along with the sequencing of the genome. Seventh, the markers would be intrinsically distributed more evenly than those based on genes.

By use of this approach, a high-density genetic map with precisely placed SNP markers that are evenly placed in the genome can be assembled with minimal effort and will be available for use to study complex genetic traits as soon as the genome sequencing is completed in year 2005 (Collins and Galas 1993).

METHODS

DNA Sequences

Three overlap regions containing SNPs were identified by the GSC for this study. On chromosome 5p15.2, an 81,830-bp overlap between BAC clones GS113H23 (GenBank accession

no. AC003015) and GS330J10 (accession no. AC002380); on chromosome 7q21-7q22, a 59,048-bp overlap between BACs RG293F11 (accession no. AC000066) and RG104F04 (accession no. AC003086); and the BRCA2 gene region on chromosome 13q12-13q13, a 59,739-bp overlap between PAC clones 257C22A (accession no. AC002525) and 96A18A (accession no. U73331).

Primary Analysis of SNPs

The GSC provided us with the ability to access the database remotely and the primary assembly data was viewed by use of the XGAP program (Bonfield et al. 1995). Up to four aligned sequencing traces could be opened and viewed simultaneously for close inspection. In the XGAP program, one can set the level of discrepancies at each nucleotide position over which it is declared an ambiguous base. When the limit is set at 80%, all differences between the consensus sequences from the two overlapping clones are designated as ambiguous and flagged. Given the fact that the base-calling and assembly programs used (PHRED, Ewing et al. 1998; PHRAP, P. Green, pers. comm.) take into account the sequence data quality, one can easily tell the sequence variations caused by poor data quality from the real polymorphisms. These polymorphisms were unmistakable because all subclones from one PAC exhibited one nucleotide but all subclones of the second PAC possessed another nucleotide. Because the primary data were available for quality check, variations caused by base-calling errors were easily eliminated. The sequence context of each polymorphism was recorded at this step. Simple sequence repeats and long runs of poly(A)s or poly(T)s were eliminated from further consideration.

Annotation and masking of common repetitive elements (such as Alu and L1 repeats) were done automatically by the GSC and these regions were removed from further consideration. PCR assays were designed for all the remaining SNPs by use of the oligonucleotide selection program (Hillier and Green 1991).

[Note: At the GSC, the sequence of only one of the two overlapping clones is fully finished and deposited to GenBank. Typically, the shotgun sequence data in the overlapping region found in the second clone sequenced are assembled into contigs with a few gaps in between. The prefinished data are archived and are not deposited in a public database. However, all of the sequencing traces for both clones are freely accessible. Any researcher interested in the overlapping sequences is encouraged to contact the GSC for access to these data.]

Determining Frequencies of SNPs in Population Pools

All PCR assays were amplified against the complete hydatidiform mole 1 (CHM1) DNA, a completely homozygous DNA described in detail previously and the Caucasian, African-American, and Hispanic population pools (30 anonymous individuals each) (Taillon-Miller et al. 1997). PCR reaction conditions and preparation of DNA template for sequencing have been described previously (Kwok et al. 1994). DNA sequencing was done with the dichloro-rhodamine dye terminators analyzed on the 377 DNA Sequencer (PE Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. The frequency of each allele in a population pool was determined by comparing the DNA sequencing trace of a PCR

product amplified from a pooled DNA sample with that of a PCR product amplified from the DNA sample of the CHM1 (Kwok et al 1994). The CHM1 DNA serves as a homozygous control, its normalized peak height equal to a frequency of 100%.

ACKNOWLEDGMENTS

We thank M. Boyce-Jacino for pooled DNA samples; and Ellen Piernot, Jenna Putzel, and Jenica Lee for technical assistance. This work is supported by the National Institute of Health and National Science Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bonfield, J.K., K.F. Smith, and R. Staden. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* 23: 4992-4999.
- Collins, F.S. and D. Galas. 1993. A new five-year plan for the U.S. Human Genome Program. *Science* 262: 43-46.
- Collins, F.S., M.S. Guyer, and A. Chakravarti. 1997. Variations on a theme: Cataloging human DNA sequence variation. *Science* 278: 1580-1581.
- Cooper, D.N., B.A. Smith, H.J. Cooke, S. Niemann, and J. Schmidtke. 1985. An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* 69: 201-205.
- Ewing, B.G., L. Hillier, M.C. Wendl, and P. Green. 1998. Basecalling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Res.* 8: 175-185.
- Hillier, L. and P. Green. 1991. OSP: A computer program for choosing PCR and DNA sequencing primers. *PCR Methods Applic.* 1: 124-128.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, UK.
- Kruglyak, L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nature Genet.* 17: 21-24.
- Kwok, P.-Y., C. Carlson, T. Yager, W. Ankener, and D.A. Nickerson. 1994. Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* 23: 138-144.
- Kwok, P.-Y., Q. Deng, H. Zakeri, and D.A. Nickerson. 1996. Increasing the information content of STS-based genome maps: Identifying polymorphisms in mapped STSs. *Genomics* 31: 123-126.
- Marra M.A., T.A. Kucaba, N.L. Dietrich, E.D. Green, B. Brownstein, R.K. Wilson, K.M. McDonald, L.W. Hillier, J.D. McPherson, R.H. Waterston. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* 7: 1072-1084.
- Marshall, E. 1997. Snipping away at genome patenting. *Science* 277: 1752-1753.
- Taillon-Miller, P., I. Bauer-Sardiña, H. Zakeri, L. Hillier, D.G. Mutch, and P.-Y. Kwok. 1997. The homozygous complete hydatidiform mole: A unique resource for genome studies. *Genomics* 46: 307-310.

Received February 26, 1998; accepted in revised form April 28, 1998.

Simple and sensitive detection of mutations in the ras proto-oncogenes using PNA-mediated PCR clamping

Christian Thiede, Ekkehard Bayerdörffer¹, Rainer Blasczyk, Burghardt Wittig² and Andreas Neubauer*

Abteilung für Innere Medizin m.S. Hämatologie und Onkologie, Virchow-Klinikum, Medizinische Fakultät der Humboldt Universität zu Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, ¹Abteilung für Innere Medizin/Gastroenterologie, Klinikum Großhadern der Ludwig-Maximilians-Universität, Marchioninistrasse 15, 81377 München, Germany and ²Abteilung Molekularbiologie und Bioinformatik, Institut für Molekularbiologie und Biochemie, Freie Universität Berlin, Arnimallee 22, 14197 Berlin, Germany

Received November 21, 1995; Revised and Accepted January 17, 1996

Point mutations in the ras proto-oncogenes are amongst the most frequent changes found in human malignancies (1) and may have prognostic importance. A variety of methods have been published for the detection of ras mutations (2-4). However, the most frequently used assays are limited either by an unsatisfactory sensitivity (2,3), the spectrum of detectable mutations (4) or require radioactive labeling (2,3). Here, we present a novel approach for the detection of ras point mutations based on the recently described method of peptide nucleic acid (PNA) mediated PCR clamping (5). PNAs are DNA-mimics where the phosphoribose backbone is replaced by a peptide-like repeat of (2-aminoethyl)-glycine units. Due to this chemical difference PNAs differ from DNA molecules in several aspects: (i) PNA/DNA-hybrids have a higher thermal stability compared with the corresponding DNA/DNA hybrids ($\sim 1^\circ\text{C}/\text{base}$ for mixed sequences); (ii) PNA/DNA hybrids are more destabilized by single base pair mismatches than the corresponding DNA/DNA hybrids (5); and (iii) PNAs could not serve as primer molecules in PCR.

The basic idea to use these features for the detection of ras gene mutations was to extend the original assay described for mutations at a single position (5) to several mutations clustered in a 4-5 bp span in one PCR. The principle was to hybridize chromosomal human DNA to a 15mer PNA complementary to the wild-type (wt) Ki-ras sequence surrounding codons 12 and 13 (schematically illustrated in Fig. 1). We reasoned that, in the case of wt Ki-ras, formation of PNA/DNA hybrids would be favoured. The bound PNA should sterically hinder annealing of a partially overlapping generic oligonucleotide, thus excluding the normal Ki-ras sequence from sufficient PCR amplification. In the case of mutant alleles, the melting temperature of the PNA/DNA hybrid was reduced, thereby allowing the 23mer oligonucleotide to outcompete PNA annealing and preferential amplification of mutant sequences.

This model was tested on six of 12 possible Ki-ras mutations in codons 12 and 13 derived from several tumor entities, which were available to us and had been characterized previously (6,7). A concentration of 2.84 μM PNA-1 was found sufficient to inhibit any detectable Ki-ras amplification starting from wt DNA (data not shown). In contrast, even at 14.2 μM PNA-1 no reduction in

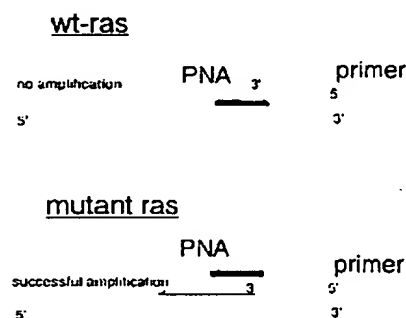


Figure 1. Schematic illustration of PNA mediated PCR clamping for the detection of Ki-ras point mutations (for details, see text).

amplification was seen with primers for the human γ -interferon gene (8) or the human growth hormone gene, indicating the specificity of this inhibition. Several reaction parameters were evaluated for their influence on discrimination between mutant and wt Ki-ras alleles, e.g. the T_m of the competing generic primer, PCR temperature profiles, the total number of cycles, buffer composition and different polymerases. Among these, the total number of PCR cycles, the oligonucleotide annealing temperature and the amount of template DNA added were most important. A significant improvement was achieved by addition of glycerol (7.5% v/v) to the reaction.

After optimization the assay was able to detect all tested Ki-ras mutations irrespective of the site or type of the mutation in a single PCR (Fig. 2). A 429 bp fragment of the human growth hormone gene was coamplified in each reaction to exclude unspecific PCR inhibition in cases with no Ki-ras amplification. We also checked the sensitivity of our method by diluting a sample carrying a heterozygous GGT \rightarrow GAT mutation in codon 12 in wt DNA. Under the conditions described, the mutation was detected down to one mutant allele in a background of 200 wt alleles. Direct sequencing confirmed that the mutant allele was amplified predominantly (Fig. 3D), whereas sequence analysis of the same sample reacted without PNA failed to detect the alteration (Fig. 3C).

* To whom correspondence should be addressed

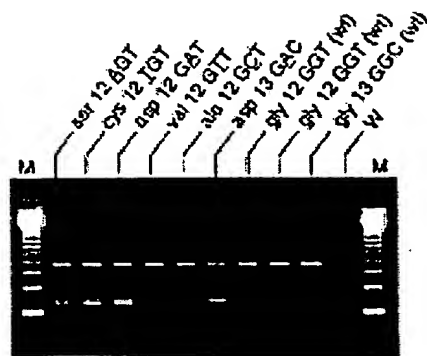


Figure 2. Detection of different Ki-ras mutations in codons 12 and 13 using PNA mediated PCR clamping. The 157 bp Ki-ras product in lanes 2–7 indicates the presence of a mutation in the corresponding tumor, whereas wt-controls (lanes 8–10, lower band) are negative for Ki-ras, even with a 3-fold excess of DNA (0.45 µg) added (lane 10). Coamplification of a 429 bp fragment of the human growth hormone gene (HGH) was done in each reaction to exclude unspecific inhibition in negative cases (lanes 2–10, upper band). PCR was performed in 50 µl, containing 100 µM deoxynucleoside-triphosphates, 0.001% gelatine, 50 mM KCl, 1.5 mM MgCl₂ and 10 mM Tris-HCl (pH 8.3), 0.25 µM KRAS-1 (5'-GTACTGGTGGAGTATTTGATAGTG-3') and KRAS-2 (5'-ATCGTCAAGGCACTCTTGCCCTAC-3') primers, 0.12 µM HGH-s (5'-GCCT-TCCCAACCAATTCCTTA-3') and HGH-as (5'-TCACGGATTCTGTGTGTGTTTC-3') primers, 2.84 µM PNA-1 (H₂N-TACGCCACCACTCC-CON₂H; Perceptive Biosystems, Freiburg, Germany), 7.5% glycerol (v/v) and 0.15 µg template DNA. The mixture was covered with 50 µl paraffin oil. To prevent unspecific polymerization prior to thermal cycling, hot start was performed by adding *Taq* polymerase (1 U; Perkin Elmer Cetus) after 5 min incubation at 94°C. PCR consisted of 28 cycles with 94°C/60 s, 70°C/50 s, 58°C/50 s and 72°C/60 s, with 180 s at 94°C in the first cycle and an additional final extension cycle with 94°C/1 min and 60°C/10 min. The additional 70°C step was performed in order to achieve preferential annealing of the PNA. Ten microliters of the reaction were electrophoresed on a 3% agarose-gel and stained with ethidium bromide (0.5 µg/ml).

In our eyes the major advantage of PNA mediated PCR clamping over published assays seems to be the higher flexibility which allows detection of mutations stretched over 4–6 bp in a single reaction. In addition, this method is not restricted to specific base exchanges, a major drawback of procedures using allele specific amplification (9). In conclusion, PNA-mediated PCR clamping is an attractive tool for the detection of ras gene point mutations. The simplicity and versatility make it especially helpful in large scale screening programs. Due to the special situation, ras gene mutations are ideally suited targets, but the assay could also be a rapid and sensitive prescreening method for common clustering mutations in other genes, such as hot-spot mutations in codons 175, 248 or 273 of the p53 tumor suppressor gene.

ACKNOWLEDGEMENT

This study was supported in part by the Deutsche Forschungsgemeinschaft through grants to A.N. (Ne 310/6-2) and B.W. (SFB 366), and by the Wilhelm-Sander Stiftung (A.N.).

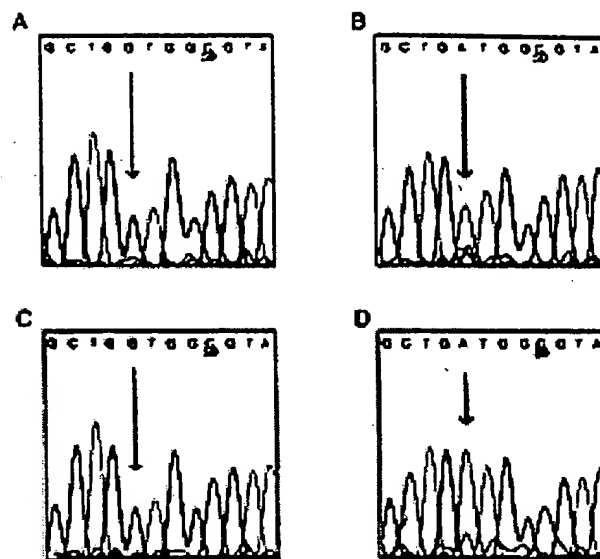


Figure 3. Sequence analysis of samples amplified with (D) or without (A, B and C) PNA reveals preferential amplification of mutant alleles mediated by the PNA. (A) Wt Ki-ras sequence with GGT in codon 12 derived human placental DNA. (B) Heterozygous GGT → GAT mutation in the second position of codon 12 derived from a liver metastasis of colorectal cancer. (C) Mutant DNA, diluted 1:10 in human placental DNA, was amplified without PNA; the mutation is not detectable. (D) The same sample reacted in the presence of PNA-1, predominantly the mutant allele could be detected. Automated, fluorescence solid-phase sequencing of PCR-products using T7-dye terminator chemistry [Applied Biosystems (ABI), Foster City, CA, USA] was performed on a 373A DNA sequencing system (ABI) essentially as recommended by the manufacturer.

REFERENCES

- 1 Bos, J.L. (1989) *Cancer Res.*, **49**, 4682–4689.
- 2 Farr, C.J., Saiki, R.K., Ehrlich, H.A., McCormick, F. and Marshall, C.J. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1629–1633.
- 3 Suzuki, Y., Orita, M., Shiralshi, M., Hayashi, K. and Sekiya, T. (1990) *Oncogene* **5**, 1037–1043.
- 4 Haliassos, A., Chomel, J.C., Grandjouan, S., Kaplan, J.C. and Kitzis, A. (1989) *Nucleic Acids Res.*, **17**, 8093–8099.
- 5 Orum, H., Nielsen, P.E., Egholm, M., Berg, R.H., Buchardt, O. and Stanley C. (1993) *Nucleic Acids Res.*, **21**, 5332–5336.
- 6 Neubauer, A., Dodge, R.K., George, S.L., Davey, F.R., Silver, R.T., Schiffer, C.A., Mayer, R.J., Ball, E.D., Wurster-Hill, D., Bloomfield, C.D. and Liu, E.T. (1994) *Blood*, **83**, 1603–1611.
- 7 Rochlitz, C.F., Heide, I., de Kant, E., Boehmer, R., Peter, F.J., Neuhaus, P., Huhn, D. and Herrmann, R. (1993) *Oncology*, **50**, 70–76.
- 8 Neubauer, A., Neubauer, B. and Liu, E.T. (1990) *Nucleic Acids Res.*, **18**, 993–998.
- 9 Takeda, S., Ichii, S. and Nakamura, Y. (1993) *Hum. Mutat.*, **2**, 112–117.

Mutation detection using immobilized mismatch binding protein (MutS)

Robert Wagner*, Paul Debbie and Miroslav Radman¹

Genecheck Inc., 400 East Horsetooth Road, Fort Collins, CO 80525, USA and ¹Institute Jacques Monod, 2 Place Jussieu, Tour 43, 75251 Paris Cedex 05, France

Received June 15, 1995; Revised and Accepted August 20, 1995

ABSTRACT

An accurate and highly sensitive mutation detection assay has been developed. The assay is based on the detection of mispaired and unpaired bases by immobilized mismatch binding protein (*Escherichia coli* MutS). The assay can detect most mismatches and all single base substitution mutations, as well as small addition or deletion mutations. The assay is simple to use and does not require the use of either radioactivity or gel electrophoresis.

INTRODUCTION

A large number of human genetic diseases are caused by small genetic alterations, including single base substitutions and small additions or deletions. Such mutations may be inherited (inherited syndromes), arise *de novo* in the germline (sporadic diseases) or be acquired somatically (e.g. cancers). The development of diagnostic tests for such small DNA alterations will facilitate both the prevention and treatment of a wide variety of diseases. In addition, the ability to scan a large number of DNA samples for small differences will be useful for large scale studies of polymorphism in human and other species and in identifying unknown genes.

The methods that, to date, have been most successful in detecting small genetic alterations fall into two broad categories: (i) those based on sequence- or mismatch-dependent variability in electrophoretic mobilities; (ii) those based on proteins capable of detecting mispaired bases in heteroduplex DNA (1-7). The former class, while reasonably accurate, is technically demanding and requires the use of polyacrylamide gel electrophoresis, thus making the assays labor intensive, somewhat difficult to automate and difficult to apply to the rapid screening of a large number of samples. Mismatch detection assays also fall into two broad classes: (i) those which involve chemical or enzymatic cleavage of mismatch-containing heteroduplexes at the site of a mismatch (1,2,5,6); (ii) those which involve binding of mismatch-containing heteroduplexes (7). All mismatch cleaving assays and most mismatch binding assays require gel electrophoresis. Mismatch detection assays involving gel mobility shifts require the identification of protein-DNA complexes in polyacrylamide gels (3). Cleavage of mismatch-containing heteroduplexes requires subsequent identification of specific fragments via gel electro-

phoresis, as do mismatch binding assays involving nuclease protection (4).

Enzymatic mismatch cleavage recognizes distortions produced by disruptions in base pairing, such that those mismatches which produce maximal helical distortion and occur in A:T-rich regions are best recognized. However, the most frequently occurring replication errors arise from mismatches (or unpaired bases) which cause minimal helical distortion and occur most frequently in G:C-rich regions (8), which may make it difficult for enzymatic cleavage methods to detect some of the most commonly occurring mutations.

The specificity of mismatch binding proteins involved in mismatch, i.e. replication error, repair *in vivo* should make them ideally suited to mutation detection. Mismatch binding proteins recognize best those mismatches and unpaired bases which most resemble base pairs and which are, therefore, most likely to occur as replication errors. In addition, mismatch binding proteins recognize mismatches best in regions of high G:C content. However, the use of mismatch binding proteins in mutation detection has, heretofore, met with limited success. The results reported here indicate that immobilized mismatch binding protein exhibits enhanced ability to discriminate between DNA with and without mismatches relative to mismatch binding protein in solution. Thus immobilization facilitates mutation detection by mismatch binding protein and represents a novel approach to mutation/polymorphism detection. The assay is simple to use, accurate and readily amenable to automation.

MATERIALS AND METHODS

Oligonucleotides

Oligonucleotides of the sequence biotin-GCACCTGACTC-CTGXGGAGAAGTCTGCCGT were annealed to unlabeled complementary oligonucleotides to form all possible mismatches (heteroduplexes) and a G:C base pair (homoduplex). Heteroduplexes were also prepared with unpaired bases by inserting the following bases between positions 15 and 16 of the complementary (non-biotinylated) strand of homoduplex molecules: (i) C, (ii) CA, (iii) CAG, (iv) CAGG.

Immobilized mismatch binding protein assay

MutS (Genecheck Inc.; 500 ng/well) in reaction buffer (20 mM Tris-HCl, pH 7.6, 5 mM MgCl₂, 0.1 mM dithiothreitol, 0.01 mM

* To whom correspondence should be addressed

EDTA) was bound to nitrocellulose pre-wet with reaction buffer in a 48-well slot blotting apparatus (Hoeffer). Reaction buffer without MutS was added to control wells. After 20 min at room temperature nitrocellulose was blocked with 200 μ l/well 3% horseradish peroxidase (HRP)-free bovine serum albumin (BSA). After 1 h excess blocking solution was removed under vacuum and DNA (1 and 10 ng) was added in 20 μ l reaction buffer plus 3% BSA. After 30 min at room temperature wells were washed five times with 100 μ l reaction buffer. All washes were poured out rather than removed under vacuum. Streptavidin-conjugated HRP (100 μ l, 0.05 μ g/ml) in reaction buffer plus BSA was added to each well. After 20 min at room temperature any remaining solution was poured out and the wells washed five times with 100 μ l reaction buffer as described above. The nitrocellulose sheet was removed from the apparatus, washed four times with 50 ml reaction buffer, blotted dry, immersed in 10 ml ECL development solution (Amersham) for 1 min, blotted dry and exposed to X-ray film.

Human genomic DNA

Human genomic DNA was PCR amplified to obtain the following specific fragments of the human glucokinase gene.

Exon 3.

Het-3a. The template was human genomic DNA known to be heterozygous for a transition mutation (G:C→A:T) in exon 3 of the glucokinase gene. The DNA was obtained from CEPH (Paris, France).

Het-3b. The template was human genomic DNA known to be heterozygous for a transversion mutation (G:C→C:G) in exon 3 of the glucokinase gene. The DNA was obtained from CEPH (Paris, France).

Hom-3. The template was human genomic DNA presumed to be homozygous in exon 3 of the glucokinase gene. The DNA was obtained from Sigma.

The primers were 5'-biotin-GGCTGACACACTTCTCTCT and 5'-GATGGAGTACATCTGGTGTT. The amplified fragment was 150 bp long.

Exon 6.

Het-6. The template was human genomic DNA known to be heterozygous for a transition mutation (G:C→A:T) in exon 6 of the glucokinase gene. The DNA was obtained from CEPH (Paris, France).

Hom-6a. The template was human genomic DNA presumed to be homozygous in exon 6 of the glucokinase gene. The DNA was obtained from Sigma.

Hom-6b. The template was human genomic DNA known to be homozygous in exon 6 of the glucokinase gene. The DNA was obtained from CEPH (Paris, France).

The primers were 5'-biotin-CAGCTTCTGTGCTTCTTG and 5'-TGAAGCCGTTTGTACACAG. The amplified fragment was 187 bp long.

Exon 2.

Het-2. The template was human genomic DNA known to be heterozygous for a transversion mutation (G:C→T:A) in exon 2 of the glucokinase gene. The DNA was obtained from CEPH (Paris, France).

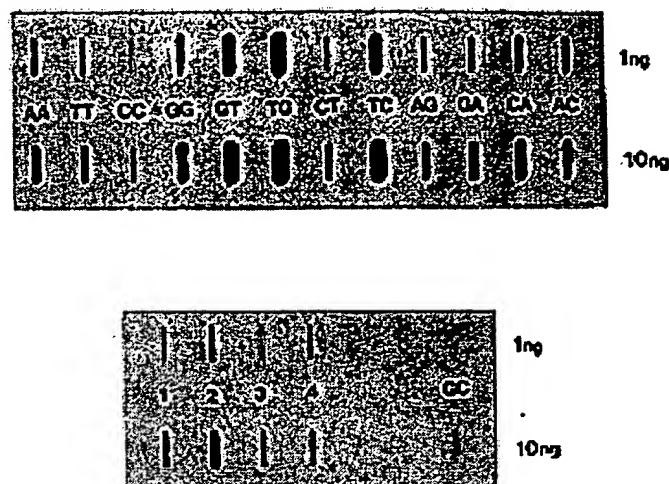


Figure 1. Binding of synthetic oligonucleotides by immobilized MutS. DNA (see Materials and Methods) with mismatches or a G:C base pair at position 15 or with one to four unpaired bases between positions 15 and 16 bound by MutS immobilized on nitrocellulose and revealed by chemiluminescence. Data are from a single experiment. Exposure time was 1 min.

Hom-2. The template was human genomic DNA presumed to be homozygous in exon 2 of the glucokinase gene. The DNA was obtained from Sigma.

The primers were 5'-biotin-GAAGGTGATGAGACGGAT and 5'-CCCAGGAGATTCTGTCTC. The amplified fragment was 230 bp long.

RESULTS AND DISCUSSION

Mismatch binding protein was immobilized by binding to nitrocellulose. The mismatch binding protein utilized in the experiments reported here is *Escherichia coli* MutS, which operates *in vivo* as the mismatch recognizing component of the *E. coli* mismatch repair system (9). Although all mismatches are not repaired with equal efficiency, either *in vivo* (10) or *in vitro* (11), MutS has been shown to bind *in vitro* to all mismatches and to heteroduplexes with one to four unpaired bases (11,12).

The results reported here are from experiments in which MutS was immobilized by binding to nitrocellulose. Other solid supports, including nylon and PVDF membranes, have been successfully employed as well (results not shown). The results of experiments utilizing synthetic 5'-biotinylated 30mers with and without mismatches or unpaired bases are shown in Figure 1. The sequence of the 30mers was taken from the β -globin gene at the region surrounding the sickle cell anemia mutation. The mismatches are at position 15 and the unpaired bases are between positions 15 and 16. Signals are generated by means of chemiluminescence. Immobilized mismatch binding protein readily detects all mismatches except C:C, which is the one mismatch which has been found to be generally refractory to repair by the *E. coli* mismatch repair system, both *in vivo* and *in vitro* (10). Heteroduplexes with one or two unpaired bases are readily detected. Heteroduplexes with three or four unpaired bases are somewhat less well detected. With immobilized MutS there is excellent discrimination between mismatched and non-mismatched oligonucleotides, the ratio of binding of G:T-containing to perfectly matched oligonucleotides

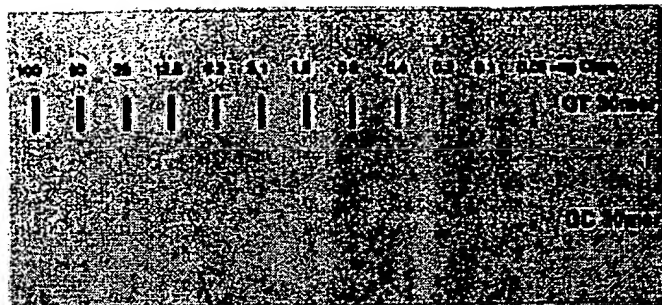


Figure 2. Binding of heteroduplex and homoduplex 30mers by immobilized MutS. DNA and assay conditions were as described in Figure 1 except that exposure time was 20 s. Data are from a single experiment.

(i.e. the ratio of the lowest concentrations at which a signal is detected) is of the order of 1000:1, whereas the ratio of binding with MutS in solution is only ~5:1 (Fig. 2 and Table 1).

Table 1. Binding of mismatch-containing DNA by MutS in solution

DNA (ng)	G:C base pair (c.p.m.)	G:T mismatch (c.p.m.)	Ratio mismatch: base pair
0.1	80	142	1.8
1.0	242	1252	5.2
10.0	1403	7236	5.2

Biotinylated oligonucleotides (described in the legend to Fig. 1) were labeled with ^{32}P by T4 polynucleotide kinase. Labeled oligonucleotides were annealed with unlabeled oligonucleotides as described in the legend to Figure 2 to produce 30mers without mismatches (homoduplexes, G:C) or heteroduplexes with G:T mismatches at position 15.

MutS (500 ng) was incubated in 20 μ l reaction buffer (20 mM Tris-HCl, pH 7.6, 5 mM MgCl₂, 0.1 mM dithiothreitol, 0.01 mM EDTA) with DNA at room temperature for 30 min. The mixtures were spotted onto 25 mm nitrocellulose filters pre-wet with reaction buffer. The filters were washed five times with 2 ml reaction buffer by vacuum filtration and dried at 80°C for 15 min. Each filter was placed in 3 ml scintillation fluid and the radioactivity determined by scintillation counting. Background counts (no MutS) were subtracted. The results presented are the means of duplicate or triplicate experiments.

The finding that T:C mismatches are better detected than C:T mismatches suggests that mismatch recognition may depend on the sequence of the individual strands, i.e. the sequence in the vicinity of the mismatch and the orientation of the mismatch with respect to the strand, at least in relatively small oligonucleotides such as those used in these experiments. However, the detectable mismatches (all mismatches except C:C) are detectable independent of orientation. In addition, G:T and T:G mismatches are detected equally well, suggesting that well-detected mismatches are well detected independent of strand orientation. It cannot be excluded that some of the variation observed in the extent of binding is due to errors in the oligonucleotide synthesizing process. Some 30mers have been observed to give signals when in a homoduplex configuration (data not shown). However, these signals are generally weaker than the weakest signals considered to be indicative of mismatch-specific binding, i.e. weaker than or equal to the C:C signal in Figure 1.

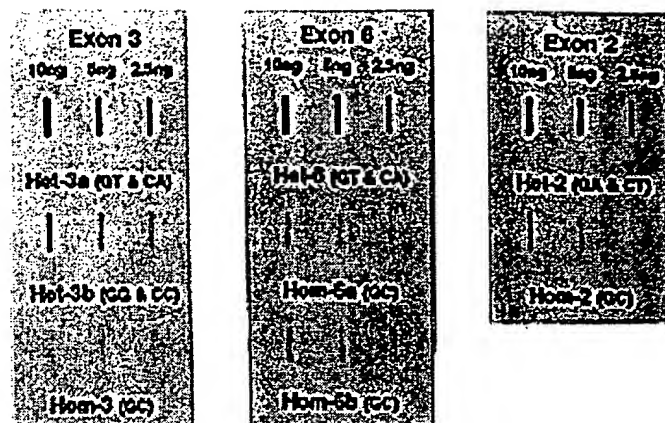


Figure 3. Detection of heterozygotes in the human glucokinase gene. Human genomic DNA from heterozygotes and homozygotes in the human glucokinase gene was amplified with primers specific to regions of exons 2, 3 and 6 (see Materials and Methods). Annealed PCR products were used in assays with immobilized MutS as described. Only DNA from heterozygotes should contain mismatches as indicated. Data from exons 3 and 6 are from a single experiment. Data from exon 2 are from a separate experiment. PCR mixture (100 µl): 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl₂, 0.2 mM dNTPs, 1.0 µM primers, 200 ng template DNA, 2.5 U Taq polymerase (Boehringer Mannheim). Thirty cycles: denaturation, 1 min at 94°C; annealing, 1 min, three cycles at 62°C, three cycles at 60°C, three cycles at 58°C, three cycles at 56°C. 18 cycles at 54°C; extension, 2 min at 72°C. Primers were removed with QIAquick spin columns (Qiagen). PCR products were eluted in 10 mM Tris-HCl, pH 8.0, adjusted to 0.1 M NaCl. Denaturation was at 100°C for 3 min. Annealing was for 1 h at 55°C, 4 min at 75°C and 30 min at 55°C. DNA was stored at 4°C until use. Exposure time 30 s.

The failure to bind C:C mismatches to a significant extent does not diminish the utility of this method for mutation detection, since every wild-type/mutant pairing gives rise to two different mismatches (e.g. G:G and C:C). G:G mismatches give strong signals.

Mutations in the human glucokinase gene are responsible for non-insulin-dependent diabetes (13). Regions of three glucokinase exons were PCR amplified from human genomic DNA known to be heterozygous for mutations in those regions and from human genomic DNA known or presumed to be homozygous for the wild-type sequence in those regions. In each case one of the primers contained a 5'-biotin, allowing detection by chemiluminescence. The same primers were used to amplify both heterozygous and homozygous genomic DNAs and the amplifications were performed simultaneously. Estimates of DNA quantities in the PCR products were obtained by polyacrylamide gel electrophoresis. (The DNA quantities shown in the figures are approximate. For the immobilized mismatch binding protein assay to produce accurate results it is sufficient to establish accurate relative quantities for homozygote and heterozygote comparisons. Any method capable of accurately obtaining such relative quantitation of the PCR products would be equally suitable.) The DNAs were denatured, by heating, allowed to re-anneal and tested for the presence of mismatches, i.e. heterozygotes, by testing their binding in an immobilized mismatch binding protein assay utilizing *E.coli* MutS.

The results are presented in Figure 3. In each case heterozygotes can be clearly distinguished from homozygotes. The actual ratios of mismatch-containing DNA binding to mismatch-free DNA binding (i.e. heteroduplex binding to homoduplex binding) are approximately twice the apparent ratios seen in Figure 3, since the

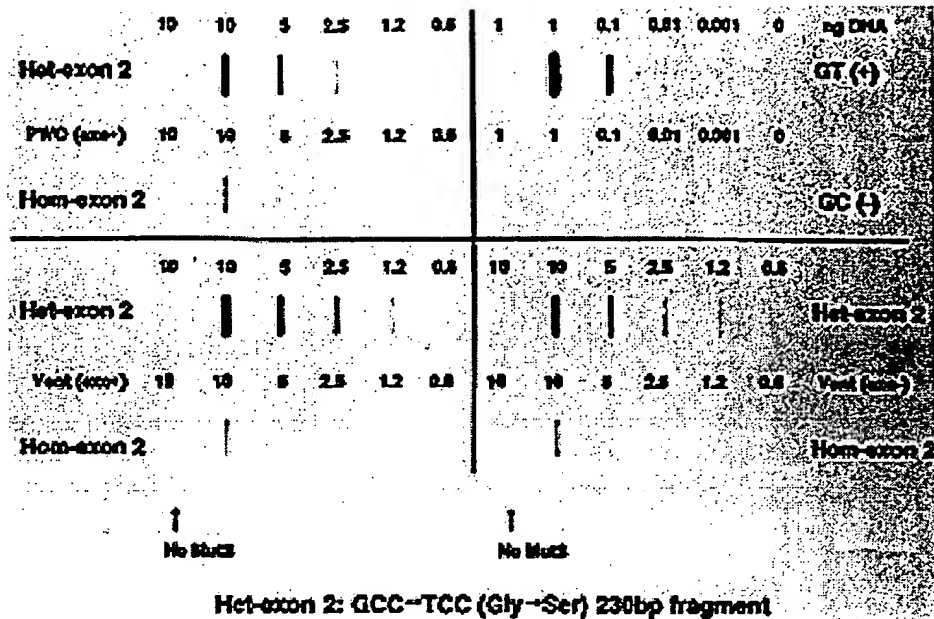


Figure 4. Comparison of products of different PCR polymerases using the immobilized MutS assay. DNA was amplified from human genomic DNAs heterozygous or homozygous in exon 2 of the glucokinase gene (see Materials and Methods). All data are from a single experiment. PCR mixtures (100 μ l): 0.25 mM dNTPs, 0.2 μ M primer 1, 0.2 μ M primer 2, 200 ng template DNA. PWO polymerase (Boehringer Mannheim): 10 mM Tris-HCl, pH 8.85, 25 mM KCl, 5 mM $(\text{NH}_4)_2\text{SO}_4$, 6 mM MgSO_4 , 5 U DNA polymerase. Vent polymerase and Vent + exonuclease: 20 mM Tris-HCl, pH 8.8, 10 mM KCl, 2 mM MgSO_4 , 10 mM $(\text{NH}_4)_2\text{SO}_4$, 0.1% Triton X-100, 2 U DNA polymerase. Thirty cycles: denaturation, 1 min at 94°C; annealing, 1 min, three cycles at 64°C, three cycles at 62°C, three cycles at 60°C, three cycles at 58°C, three cycles at 56°C, 15 cycles at 54°C; extension, 2 min at 72°C. Primers were removed with QIAquick spin columns (Qiagen). PCR products were eluted in 10 mM Tris-HCl, pH 8.0, adjusted to 0.1 M NaCl. Denaturation was at 100°C for 2 min. Annealing was for 1 h at 55°C, 4 min at 75°C and 30 min at 55°C. Exposure time 30 s.

heterozygote samples were randomly annealed. Therefore, half the molecules will be heteroduplexes and half will be homoduplexes. The strength of the heterozygote signal appears to be mismatch-dependent. In the case of exon 3, where two different mismatch pairs were studied, a strong signal is observed when the mismatches formed are G:T and C:A (Het-3a), whereas a somewhat weaker signal is observed with G:G and C:C mismatches (Het-3b), presumably due to the fact that only G:G mismatches are detected. The intermediate strength signal observed with the exon 2 fragment (Het-2) may reflect mismatch specificity, i.e. G:A and C:T mismatches appear to be somewhat less well recognized than G:T and A:C mismatches. However, the signal may also be somewhat lower because the molar concentration of mismatches is lower in the exon 2 fragment experiment than in the exon 3 fragment experiment, i.e. equal quantities of DNA were used and the fragments differ in length (230 versus 150 bp, respectively).

There is significantly increased binding of homoduplex DNA in these experiments relative to those with 30mer oligonucleotides (Fig. 2). It may be that the biotinylated primers occasionally initiate replication at sites other than the selected site. These fragments would be labeled and might be bound by immobilized MutS, either because they form mismatches when annealed with the genomic DNA from the homologous chromosome or because they form some secondary structure with mismatches. Alternatively, the homoduplex binding may be the result of polymerase errors or DNA damage occurring during amplification. Polymerase errors would be expected to occur relatively randomly throughout the amplified fragment, such that they would not be

detectable by sequencing, but the cumulative effect of such errors could be to produce a sizable fraction of PCR products with some error. These would generally produce mismatches when denatured and annealed and thus contribute to positive signals in the immobilized mismatch binding protein assay. However, when the exon 2 fragment is amplified by four different polymerases, some of which have increased fidelity of replication and should, therefore, have a reduced rate of production of error-containing fragments, the ratio of heteroduplex to homoduplex binding does not change significantly (Figs 3 and 4).

The results presented here are concerned only with the detection of heterozygous mutations. The detection of homozygous mutations can easily be accomplished by adding known homozygous DNA to the test DNA before denaturation and annealing, either before or after amplification. Thus the use of immobilized mismatch binding protein assays for mismatch, mutation, heterozygosity or polymorphism detection involving single base substitutions and small additions or deletions seems to be limited only by the need to provide substrates free of labeled DNA with random mismatches, as discussed above. Immobilized mismatch binding protein provides a simple, accurate and easy to automate system for the following.

(i) Diagnostic screening for any disease causing mutation (or mutations), including single base substitutions and small additions or deletions, for which the sequence and location of the mutation(s) are known. It is possible to detect both carriers (heterozygotes) and affected patients (homozygotes) and to distinguish between them.

(ii) Rapid and large scale screening of human (or other) genomic DNA for single base change or small addition/deletion polymorphisms. The ease and speed of the system make it

possible to screen large numbers of individuals and to construct high resolution maps based on genomic polymorphism.

In addition, it may be possible to use immobilized mismatch binding protein to remove error-containing molecules from PCR samples, to bind heterozygous sequences to allow determination of identity by descent and to study closely related varieties and/or species to characterize biodiversity.

ACKNOWLEDGEMENTS

We thank Drs Nadine Cohen and Christine Bellanné of CEPH, Paris, France, for providing human genomic DNA and PCR primer sequences.

REFERENCES

- 1 Cotton, R.G.H. (1993) *Mutat. Res.*, **285**, 125–144.
- 2 Grompe, M. (1993) *Nature Genet.*, **5**, 111–116.
- 3 Lishanski, A., Ostrander, E.A. and Rine, J. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 2674–2678.
- 4 Ellis, L.A., Taylor, G.R., Banks, R. and Baumberg, S. (1994) *Nucleic Acids Res.*, **22**, 2710–2711.
- 5 Lu, A.-L. and Hsu, L.-C. (1992) *Genomics*, **14**, 249–255.
- 6 Youil, R., Kemper, B.W. and Cotten, R.G.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 87–91.
- 7 Wagner, R. and Radman, M. (1995) In *Methods: A Companion to Methods in Enzymology*, **7**, 199–203.
- 8 Fazakerly, G.V., Quidnard, E., Woisard, A., Guschlbauer, W., van der Marel, G.A., van Boom, J.H., Jones, M. and Radman, M. (1986) *EMBO J.*, **5**, 3697–3703.
- 9 Su, S.S. and Modrich, P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5057–5061.
- 10 Modrich, P. (1991) *Annu. Rev. Genet.*, **25**, 229–253.
- 11 Su, S.S., Lahue, R.S., Au, K.G. and Modrich, P. (1988) *J. Biol. Chem.*, **263**, 6829–6835.
- 12 Parker, B. and Marinus, M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1730–1735.
- 13 Stoffel, M., Frogule, P., Takeda, J., Zoulai, H., Vionnet, N., Nishi, S., Weber, I.T., Harrison, R.W., Pilgis, S.J., Lesage, S. et al. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7698–7702.

C24

Hybridization of synthetic oligodeoxyribonucleotides to Φ x 174 DNA: the effect of single base pair mismatch

R. Bruce Wallace, J. Shaffer*, R. F. Murphy* J. Bonner*, T. Hirose and K. Itakura

Division of Biology, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, CA 91010, and *Division of Biology, California Institute of Technology, 1201 East California Street, Pasadena, CA 91125, USA

Received 16 April 1979

ABSTRACT

Oligodeoxyribonucleotides complementary to the DNA of the wild type (wt) bacteriophage Φ x174 have been synthesized by the phosphotriester method. The oligomers, 11, 14, and 17 bases long, are complementary to the region of the DNA which accounts for the am-3 point mutation. When hybridized to am-3 DNA, the oligonucleotides form duplexes with a single base pair mismatch. The thermal stability of the duplexes formed between wt and am-3 DNAs has been measured. The am-3 DNA:oligomer duplexes dissociate at a temperature about 10°C lower than the corresponding wt DNA:oligomer duplexes. This dramatic decrease in thermal stability due to a single mismatch makes it possible to eliminate the formation of the mismatched duplexes by the appropriate choice of hybridization temperature. These results are discussed with respect to the use of oligonucleotides as probes for the isolation of specific cloned DNA sequences.

INTRODUCTION

Specific nucleic acid sequences such as rRNAs, tRNAs, mRNAs, complementary DNA, or cloned DNA sequences have proven invaluable as biological probes when radioactively labeled *in vivo* or *in vitro*. These probes have been used in studies of nucleic acid complexity (see 1 for review), measurement of specific gene frequency (2), examination of sequence divergence (3), studies on transcription of specific genes (1), and most recently in detection and isolation of cloned gene sequences (4,5,6). Of the probes presently available, those for specific single copy genes are receiving the most attention. Unfortunately, only a handful of these probes are currently available. For the most part, single copy gene probes are derived from specific highly abundant mRNAs such as globin (6), ovalbumin (5), immunoglobulins (4) and others. At present, probes for single copy gene se-

quences whose mRNAs are not abundant are very difficult to isolate.

One promising alternative to the use of isolated naturally occurring nucleic acid probes, is the use of chemically synthesized oligodeoxyribonucleotide sequences. Recently, a specific 13 nucleotide sequence complementary to the gene for yeast iso-1-cytochrome c was used to detect and isolate the gene sequence cloned in a bacteriophage λ vector (7). This was possible because the actual nucleotide sequence of a region of the gene could be deduced from genetic information. As a more general approach to the use of oligodeoxyribonucleotides as probes, we propose to use a chemically synthesized mixture of oligonucleotides whose sequences represent all possible codon combinations predicted from a particular peptide sequence within a protein. One of this mixture must be complementary to a region of DNA coding for the protein. Stringent hybridization criteria would then be used to select the single correct sequence from the mixture. As a preliminary investigation, we have chosen to study the hybridization behavior of three oligonucleotides, 11, 14, and 17 bases long, to DNA from wild type (wt) and am-3 bacteriophage $\phi\chi 174$. The three oligonucleotide sequences are complementary to wt DNA at the region encompassing the am-3 point mutation (8). Duplexes formed between the oligonucleotides and am-3 DNA contain a single mismatched base pair. This system represents a useful model for the study of the effect of mismatched base pairs on duplex formation and stability.

MATERIALS AND METHODS

Synthesis of Oligodeoxyribonucleotides

The oligodeoxyribonucleotides were synthesized by the modified triester method (9). Their use in synthetic DNA directed base change of $\phi\chi 174$ DNA has been described previously (10). The oligonucleotides were gifts of Genentech, Inc., San Francisco.

Preparation of Phage DNA

The $\phi\chi 174$ wt and am-3 DNAs were gifts from Dr. Aharon Razin. The DNAs were isolated from purified phage as described (11).

Labeling of Oligonucleotides

The synthesis of oligonucleotides leaved 5'OH. The oligonucleotides are labeled by transferring the ^{32}P from $[\gamma\text{-}^{32}\text{P}]\text{ATP}$

(1000 Ci/mole) with polynucleotide kinase (Boehringer Mannheim) as described (12). Oligonucleotide (0.2 μ g) was labeled and separated from $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ by chromatography on Sephadex G-50. The excluded peak was pooled and used directly in hybridization experiments.

Preparation and Hybridization of DNA filters

The DNA was incubated in 0.2 N NaOH at 37°C for 30 minutes, neutralized with HCl and brought to 6 X SSC at 0°C. Nitrocellulose filters (Sartorius Membranfilter, 2.5 cm, pore size 0.45 μ) were wet in H_2O , washed in 6 X SSC (1 X SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.2) and placed on a filter apparatus. The amount of DNA loaded on the filters depended on the nature of the experiment. For the thermal denaturation experiments, 0.05 μ g of wt and am-3 DNA were applied to each filter. For kinetic experiments, 0.005 μ g wt DNA and 0.025 μ g am-3 DNA were applied to each of ten filters. After application of DNA, the filters were baked at 80°C *in vacuo* for 4 hours.

For hybridization, the filters were placed in 6 X SSC, 10 X Denhardt's [10 X Denhardt's = 0.2% bovine serum albumin (Sigma), 0.2% polyvinylpyrrolidone (Sigma), 0.2% Ficoll (Sigma) (13)] at room temperature for 15 minutes. The solution was drained and replaced with 2 ml of hybridization solution (6 X SSC, 10 X Denhardt's, 0.002 μ g/ml ^{32}P labeled oligonucleotide). Unless otherwise stated, hybridization was performed at 12°C for 16 hours. The filters were then washed with multiple changes of 6 X SSC at 12°C, until no more radioactivity eluted.

Thermal Denaturation

Filters which had been hybridized and thoroughly washed as described above were used for thermal denaturation studies as follows: 5 ml of 6 X SSC was placed over the filter and the temperature raised to a specific point. Once the temperature had been reached, the filter was kept at that temperature for one minute, the 6 X SSC was then removed for measurement of the radioactivity eluted (in Aquasol 2, New England Nuclear). An additional 5 ml of 6 X SSC was then added and the procedure repeated until the desired maximum temperature was reached.

The radioactivity eluted at each temperature was integrated to determine the fraction of the duplex denatured as a function

of temperature. In order to determine the temperature at which one half of the oligonucleotide dissociates from the filter (T_d), this data was fit to the error function by a nonlinear least squares fitting program as described previously (14,15).

In each experiment, essentially all of the bound oligonucleotide was removed by the highest temperature wash. In addition, none of the ϕ X174 DNA bound to the filter was lost during the procedure, since an equal amount of ^{32}P -labeled oligonucleotide will hybridize to the same filter in a second or third experiment.

Agarose Gel Electrophoresis

Vertical agarose gels (SeaKem) were used to separate DNA samples for transfer to nitrocellulose. The gels were 15 cm X 15 cm X 0.2 cm and were electrophoresed at 200 volts for 2 hours. Undigested DNA was separated on 1% agarose gels while Hae III digested ϕ X174 am-3 RFI DNA (Bethesda Research Laboratories) was separated on 2% agarose gels. After electrophoresis, gels were stained with 0.1 $\mu\text{g}/\text{ml}$ ethidium bromide (Calbiochem) for 30 minutes and photographed through an orange filter under an ultraviolet light source. The DNA was denatured by soaking the gel in 0.4 N NaOH, 0.8 M NaCl for 30 minutes. The gel was neutralized in 0.5 M tris-HCl, pH 7.6, 1.5 M NaCl for 30 minutes in the cold and the DNA transferred to nitrocellulose filters (Millipore HAWP 00010) as described by Southern (16). The filters were baked at 80°C *in vacuo* for at least 4 hours.

The filters were hybridized in 6 X SSC, 10 X Denhardt's, 0.002 $\mu\text{g}/\text{ml}$ ^{32}P -labeled oligonucleotide at the temperatures specified in each experiment. After hybridization the filters were washed in 500 ml 6 X SSC at 12°C, blotted dry and autoradiographed using pre-flashed Kodak XR-2 X-ray film exposed between 2 intensifier screens (Cronex Lightning Plus, Dupont) at -80°C for 12-36 hours.

RESULTS

Thermal Stability of Oligonucleotide - ϕ X174 DNA Duplexes

In order to study the hybridization of synthetic oligodeoxyribonucleotides to natural DNA, we synthesized three oligo-

temperature at which
from the filter (T_d),
a nonlinear least
squares (14,15).
the bound oligo-
nucleotide wash. In addi-
tion, the filter was lost during
the labeled oligonucleo-
nucleotide second or third ex-

periments to separate DNA
gels were 15 cm X
100 volts for 2 hours.
while Hae III
(Boehringer-Mann Laboratories)
was used for electrophoresis,
gels were stained with ethan-
ol (10%) for 30 min-
utes and then under an ultra-
violet light soaking the gel
in a neutral-
ized buffer for 30 minutes in the
presence of filters (Milli-
pore). The filters
were then
10 X Denhardt's,
at temperatures
above the filters
dried and autorad-
iographically exposed be-
hind a Dupont) at

nucleotides of chain length 11, 14, and 17 which are complemen-
tary to the single stranded DNA (+ strand) of the wild type (wt)
bacteriophage $\phi\chi 174$.

The 11 mer and 14 mer are synthetic intermediates of the 17
mer. The 17 mer is complementary to nucleotides 575 through 591
in the linear sequence of $\phi\chi 174$ DNA reported by Sanger and co-
workers (8) (Figure 1). These sequences represent useful models
for the study of single base pair mismatch since duplexes formed
with am-3 $\phi\chi 174$ DNA contain one A-C base pair [amber mutation is
a G→A transition at position 587 of the DNA sequence (Figure 1)].

For hybridization of ^{32}P -labeled oligonucleotides to wt and
am-3 $\phi\chi 174$ DNA, the phage DNA was immobilized on nitrocellulose
filters. Initially, hybridizations were performed at 12°C in 6
X SSC, 10 X Denhardt's ($[\text{Na}^+] = 1.2 \text{ M}$) (see Materials and Meth-
ods). From previously published results (17-20), the duplexes
formed between the 11 mer, 14 mer or 17 mer and wt DNA were
expected to be rather stable with T_m 's greater than 30°C. Under
the conditions of the hybridization or the subsequent washing of
the filters (in 6 X SSC), the oligonucleotides do not adhere

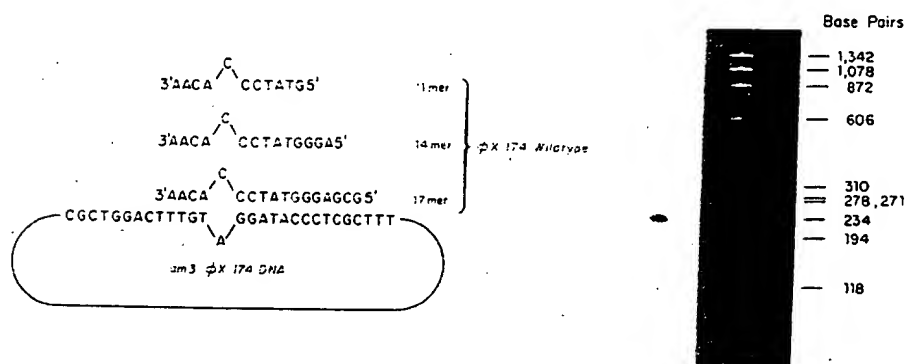


FIGURE 1. A representation of the mismatched duplexes formed between the three oligonucleotides and am-3 $\phi\chi 174$ DNA. On the right, am-3 $\phi\chi 174$ RFI DNA, which was digested with Hae III, was electrophoresed on a 2% agarose gel and blotted onto nitrocellulose as described in Materials and Methods. The filter was hybridized to ^{32}P -labeled 14 mer at 12°C, washed at 12°C and autoradiographed. It can be seen that hybridization is to the 234 base pair long restriction fragment which contains the am-3 mutation at nucleotide 587 (8).

nonspecifically to the nitrocellulose (see Figure 1).

The hybridization of all 3 oligonucleotides to wt $\phi\chi 174$ DNA was quite efficient. Between 13 and 22% of the sites on the phage DNA molecules hybridize with the labeled oligomers (Table 1). The stability of the oligonucleotide-wt DNA duplexes were examined by thermal denaturation. Filters which had been hybridized and washed at 12°C were heated to various temperatures in 6 X SSC and the radioactivity which eluted was measured. The thermal denaturation profiles are presented in Figure 2. The data is summarized in Table 1. Note that the parameter T_d , the

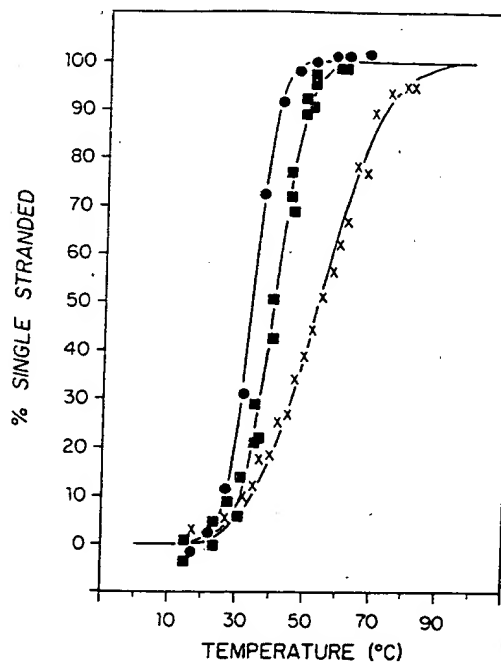


FIGURE 2. Thermal denaturation of oligonucleotide-wild type $\phi\chi 174$ DNA duplexes. The 11 nucleotide (\bullet), the 14 nucleotide (\blacksquare), and 17 nucleotide (\times) probes shown in Figure 1 were labeled with [^{32}P] in the 5' end and hybridized to wild type $\phi\chi 174$ DNA immobilized on nitrocellulose filters. The hybridization was performed in 6 X SSC, 10 X Denhardt's (13) and 0.002 μg probe/ml. The filters were washed in 6 X SSC and subjected to thermal denaturation. The radioactivity eluted at each temperature was measured and is plotted as the fraction of the total probe becoming single stranded at each temperature.

TABLE 1.

EFFECT OF MISMATCH ON HYBRIDIZATION OF OLIGONUCLEOTIDES PROBES TO ϕ X174 DNA.

Number Nucleotides In Probe	ϕ X174 DNA Hybridized	Number Nucleotides In Duplex	% Sites Hybridized	% G+C In Duplex	T_d^a Observed ^b (C°)
11	wt	11	20	46	33.2 \pm 0.4
11	am-3	10	-	-	-
14	25	14	13.6	50	40.6 \pm 0.7
14	am-3	13	2.9	43	31.1 \pm 0.8
17	wt	17	22.6	59	55.1 \pm 0.8
17	am-3	16	11.3	53	43.5 \pm 0.3

a T_d is the temperature at which one half of the duplexes are dissociated under the conditions of the experiments.

b The error estimates represent the error in the T_d parameter as calculated by the least squares fitting program.

temperature at which one half of the duplexes are dissociated, is used rather than T_m since the experiment does not allow direct measurement of T_m in a thermodynamically rigorous way. As expected, an increase in thermal stability is seen with an increase in duplex length.

Compared to the wt DNA, hybridization of the 3 oligonucleotides to am-3 DNA is much less efficient (Table 1). In fact, the level of hybridization of the 11 mer to am-3 DNA was barely above background and determination of an accurate T_d was not possible. The thermal denaturation profiles of the oligonucleotide-am-3 DNA duplexes are presented in Figure 3. The oligonucleotide-wt DNA melts are plotted for comparison. The data is summarized in Table 1. It can be seen that the thermal stability of the 14 mer and 17 mer duplexes with am-3 DNA is much lower than that of the corresponding wt duplexes.

The substantial difference in the thermal stability of perfectly matched and mismatch duplexes suggests that hybridization of the oligonucleotides to am-3 DNA could be eliminated with little, if any, effect on hybridization to wt DNA by the appropriate choice of either filter wash temperature or hybridization temperature. To test this prediction, wt and am-3 DNAs were electrophoresed in adjacent lanes of a 1% agarose gel. The DNA in the gel was transferred to nitrocellulose essentially as described by Southern (16). The filter strips were hybridized at

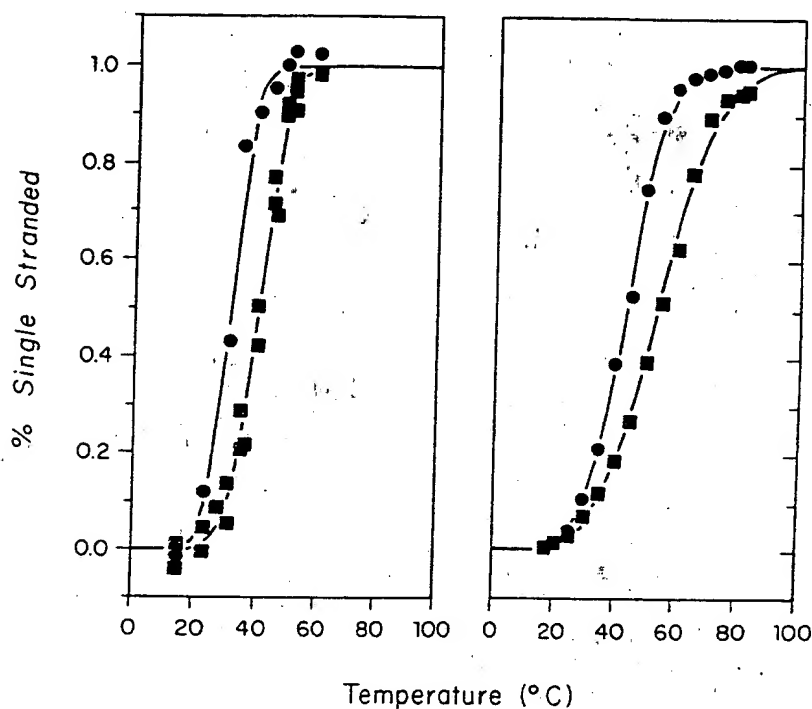
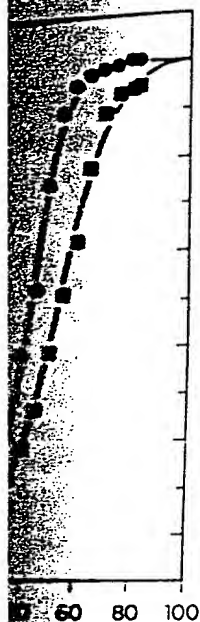


FIGURE 3. Thermal denaturation of oligonucleotide-wild type $\phi\chi 174$ DNA (■) and oligonucleotide-am-3 $\phi\chi 174$ DNA (●) duplexes. The hybridization and thermal denaturation were performed as in Figure 2. Left 14-mer and right 17-mer.

12°C with ^{32}P -labeled 11 mer, 14 mer or 17 mer, washed at 12°C and autoradiographed. Figure 4a shows the autoradiograph obtained. Hybridization to wt DNA is approximately equal for all 3 oligonucleotides. The level of hybridization to am-3 DNA, on the other hand, is dependent on the length of the oligonucleotide. Very little hybridization to am-3 DNA is seen for the 11 mer, a slightly greater amount for the 14 mer, while hybridization of the 17 mer to am-3 DNA approaches that to wt DNA (see Table 1 for comparison). After autoradiography, the filter strips were rewashed at higher temperatures, 30°C for the 11 mer, 37°C for the 14 mer, and 50°C for the 17 mer. The autoradiograph of the rewashed filters is shown in Figure 4b. It can be seen



side-wild type
DNA (o) duplexes.
performed as in

washed at 12°C
radiograph ob-
served equal for all
to am-3 DNA, on
the oligonucleo-
seen for the 11
while hybridiza-
to wt DNA (see
the filter
for the 11mer,
autoradiograph
it can be seen

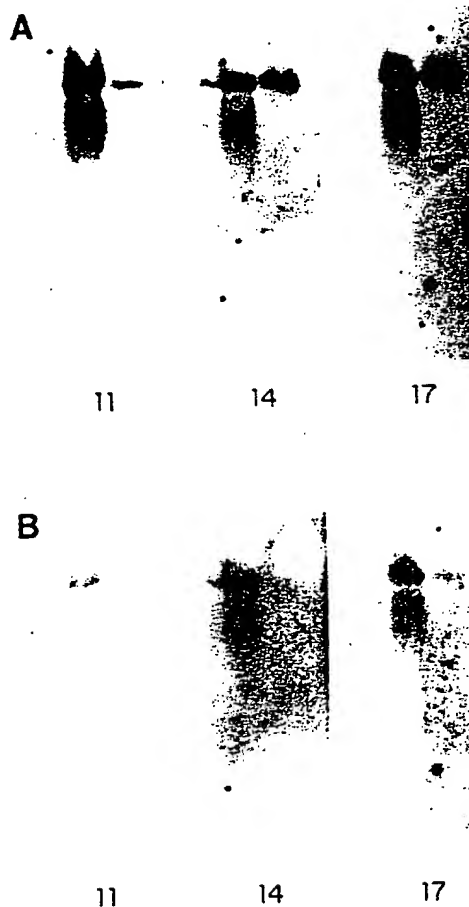


FIGURE 4. Effect of filter wash temperature on hybridization of ^{32}P -labeled oligonucleotides to wt and am-3 $\phi\text{X}174$ DNA. Equal amounts of single stranded wt (on the left) and am-3 DNA (on the right) were electrophoresed on 1% agarose gels and blotted onto nitrocellulose filters as described in Materials and Methods. Three filters, each containing one band of wt and one of am-3 DNA, were hybridized with ^{32}P -labeled 11-mer, 14-mer, and 17-mer at 12°C. The filters were washed at 12°C and autoradiographed overnight with preflashed X-ray film between two screens at -80°C (A). The filters were then rewashed at 30°C (11-mer), 37°C (14-mer) and 50°C (17-mer) and re-autoradiographed (B).

that hybridization to the am-3 DNA is virtually eliminated in each case while hybridization to wt DNA is only diminished

slightly.

Oligonucleotide Hybridization at Elevated Temperatures

In order to examine the effect of hybridization temperature on the formation of non-mismatched or mismatched duplexes, wt and am-3 DNAs were electrophoresed in alternate lanes of a 1% agarose gel and the DNA transferred to a nitrocellulose filter. The filter was cut into strips containing one band of wt and one of am-3 DNA each. The strips were hybridized in 6 X SSC, 10 X Denhardt's, 0.002 μ g 32 P-labeled 14 mer/ml at 12°C, 25°C, 30°C, 35°C, and 40°C, washed at 12°C, and autoradiographed. The results are shown in Figure 5. It can be seen that hybridization to am-3 DNA is dramatically reduced at 25°C and higher. Hybridization to wt DNA is not affected between 25°C and 35°C with only a slight decrease in hybridization at 40°C (1°C below

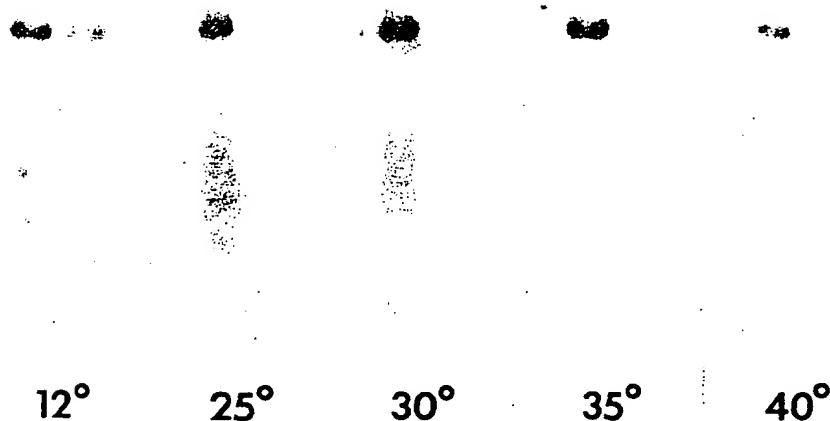


FIGURE 5. Effect of hybridization temperature on the formation of duplexes between 32 P-labeled 14-mer and wt (on the left) and am-3 DNA (on the right). Equal amounts of wt and am-3 DNAs were electrophoresed in alternate lanes of a 1% agarose gel and blotted onto nitrocellulose filters. The filters were cut into strips and hybridized at various temperatures (12°, 25°, 30°, 35° and 40°C) as described in Materials and Methods. The strips were then removed from the hybridization solution, washed briefly at 12°C and autoradiographed.

T_d).Kinetics of Oligonucleotide Hybridization

In order to determine the optimum time required for oligonucleotide hybridization, the kinetics of duplex formation were measured. Multiple filters containing wt or am-3 DNA were prepared. Individual filters were hybridized for various time, washed immediately to remove unbound probe and the radioactivity measured. Figure 6 shows the kinetics of hybridization of ³²p-labeled 14 mer to wt and am-3 DNA at 12°C and wt DNA at 37°C. The data is plotted as equivalent C_{ot} versus fraction hybridized. The data has been fit to a first order rate equation (14). The rate constants and C_{ot} 1/2 are presented in Table 2. The measured rate constant is within a factor of 3 (for 37°C) that calculated by the equation of Wetmur and Davidson (22).

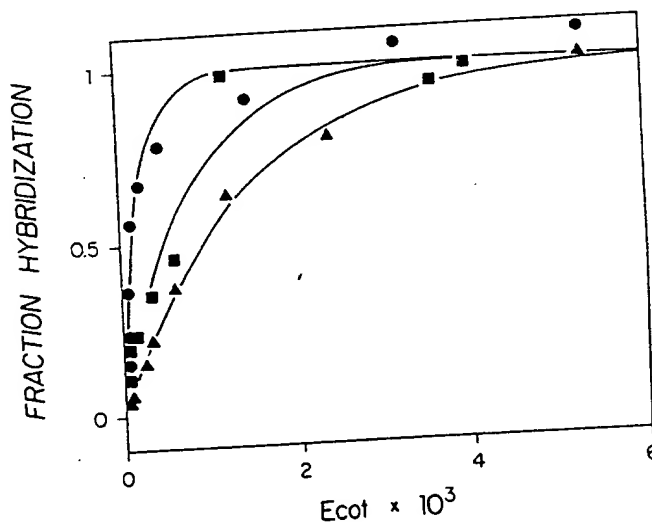


FIGURE 6. Kinetics of hybridization of ³²p-labeled 14-mer to wt DNA at 12°C (■) and 37°C (●) and to am-3 DNA at 12°C (▲). Individual filters containing 0.005 µg wt DNA or 0.025 µg am-3 DNA were hybridized for various times in 6 X SSC (1.2 M Na⁺), the filters washed, and the radioactivity determined. The C_{ot} was corrected for hybridization in 0.18 M Na⁺ as described (21). The data was fit to a first order rate equation by a non-linear least-square fitting program as described (14,15).

TO ϕ X174 DNA.

k_2 (mole
l⁻¹sec⁻¹)

5×10^{-4}
 7.2×10^{-4}
 6×10^{-4}
 6×10^{-5}

reaction as de-

has been cor-

Wetmur and
is described (21).

usually occurring
am-3 point muta-
pair mismatch ex-
an A-C base pair
other by 6 to
lined the effects
and copolymers
ned the effects
between oligo-
A sequences. It
the effects of
g DNA in order
in of mismatched
re necessary in
hybridization as

ed the use of an
a probe to iso-
13 nucleotide
the cloned cyto-
culties which
oducibility of
riment to ex-

periment and the poor signal obtained with the labeled probe in the plaque screening assay. Model studies such as this one are very useful, therefore, in establishing conditions which optimize signal and minimize nonspecific interactions.

The single base pair mismatch in the duplexes formed between the 11-, 14- and 17-nucleotide oligomers and am-3 ϕ X174 DNA has a significant destabilizing effect. The single base pair mismatch in the case of the 17 mer and the 14 mer resulted in complexes with a thermal stability approximately equivalent to a non-mismatched duplex three base pairs shorter (Table 1). A similar degree of destabilization has been reported for various single and multiple mismatched base pairs in oligonucleotide duplexes by Dodgson and Wells (26). The lower thermal stability of the am-3 DNA:oligomer hybrids make it possible to differentiate between the mismatched and non-mismatched duplexes by the appropriate choice of filter wash temperature (Figure 4) or hybridization temperature (Figure 5). These results demonstrate the high degree of specificity which can be achieved using oligonucleotide probes.

In addition to the effects on thermal stability, a single base pair mismatch has a significant effect on the level of hybridization achieved. This is most clearly seen in the autoradiograph shown in Figure 4. At lower temperature (12°C), the level of hybridization of oligomer to wt DNA is approximately equal for all three (11 to 17) while the level of hybridization to am-3 is dependent on the length of the oligomer. The reason for this phenomenon is not known, however, an examination of the rates of hybridization of the 14 mer to wt and am-3 DNAs (Figure 6) demonstrate that the difference in the levels of hybridization are not due to a lower rate of formation of the mismatched duplexes (at low temperatures). It is conceivable that the rate of dissociation of the mismatched duplexes is greater than those of perfectly matched duplexes and that this is responsible for the lower levels of hybridization to am-3 DNA over wt DNA.

ACKNOWLEDGEMENTS

This work was supported by USPHS grants to James Bonner

TABLE 2.
KINETICS OF HYBRIDIZATION OF ^{32}P -LABELED 14-MER TO ϕX174 DNA.

ϕX174 DNA	Temp. (°C)	k^a ($\text{M}^{-1}\cdot\text{sec}^{-1}$)	$\text{Cot}_{1/2}^b$ (mole nucleotide $\cdot\text{L}^{-1}\cdot\text{sec}^{-1}$)
wt	12	1251 ± 321	5.5×10^{-4}
wt	37	4037 ± 960	1.72×10^{-4}
am-3	12	722 ± 79	9.6×10^{-4}
Calculated ^c kinetic parameters		10440	6.6×10^{-5}

a Rate constant calculated for a first order reaction as described (14,15).

b $\text{Cot}_{1/2}$ is calculated from k ($\ln 2/k$). All data has been corrected for hybridization in 0.18 M Na^+ (21).

c Rate constant is calculated as described by Wetmur and Davidson (22) and corrected for 0.18 M Na^+ as described (21).

DISCUSSION

This study has taken advantage of the naturally occurring base substitution in ϕX174 DNA resulting in the am-3 point mutation (8). As such, the effects of single base pair mismatch examined here refer only to this single example, an A-C base pair flanked on one side by four base pairs and on the other by 6 to 12 base pairs. Many previous studies have examined the effects of mismatch in model RNA (23,24) and DNA homo- and copolymers (25,26). Very few studies, however, have examined the effects of single base pair mismatch on duplexes formed between oligodeoxyribonucleotides and naturally occurring DNA sequences. It was the main purpose of this study to examine the effects of such mismatches for a defined naturally occurring DNA in order to establish conditions under which the formation of mismatched duplexes could be eliminated. Such conditions are necessary in order to use the specificity of oligonucleotide hybridization as a probe for defined DNA sequences.

Montgomery, *et al.* (7) have recently reported the use of an oligodeoxyribonucleotide of defined sequence as a probe to isolate the yeast gene for cytochrome c. While the 13 nucleotide probe was used successfully in the isolation of the cloned cytochrome c gene, the authors pointed out many difficulties which were encountered. Among these were the poor reproducibility of hybridization of the ^{32}P -labeled 13 mer from experiment to ex-

Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII

RIMA YOUSSEF[†], BORRIES W. KEMPER[‡], AND RICHARD G. H. COTTON[†]

[†]Olive Miller Laboratory, Murdoch Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia; and [‡]Institute for Genetics of the University of Cologne, Weyertal 121, 50931 Cologne 41, Germany

Communicated by Leonard S. Lerman, Massachusetts Institute of Technology, Cambridge, MA, June 27, 1994 (received for review December 10, 1993)

ABSTRACT Each of four possible sets of mismatches (G·A/C·T, C·C/G·G, A·A/T·T, and C·A/G·T) containing the 8 possible single-base-pair mismatches derived from isolated mutations were examined to test the ability of T4 endonuclease VII to consistently detect mismatches in heteroduplexes. At least two examples of each set of mismatches were studied for cleavage in the complementary pairs of heteroduplexes formed between normal and mutant DNA. Four deletion mutations were also included in this study. The various PCR-derived products used in the formation of heteroduplexes ranged from 133 to 1502 bp. At least one example of each set showed cleavage of at least one strand containing a mismatch. Cleavage of at least one strand of the pairs of heteroduplexes occurred in 17 of the 18 known single-base-pair mutations tested, with an A·A/T·T set not being cleaved in any mismatched strand. We propose that this method may be effective in detecting and positioning almost all mutational changes when DNA is screened for mutations.

The detection of mutations is important, particularly in the diagnosis of inherited diseases. Changes in the DNA sequences of a gene can be harmful and it is important in our understanding of human genetics that we are able to identify and classify these alterations and the phenotypic changes that they induce. Consequently, the need for a reliable method for the detection of mutations in DNA to avoid repetitive sequencing of kilobase lengths of DNA has led to the development of a number of different screening methods that have both positive and negative attributes (see ref. 1 for a review of current mutation detection methods). Thus, the search for a reliable and efficient approach to the detection of known and unknown mutations continues.

The resolvases are an important group of enzymes that are responsible for catalyzing the resolution of branched DNA intermediates that form during genetic recombination. Their mode of action is directed by bends, kinks, or DNA deviations. These enzymes have their effect close to the actual site of DNA distortion (2). T4 endonuclease VII, the product of gene 49 of the bacteriophage T4 (2), is a resolvase that has been well characterized (3-5). It was the first enzyme shown to resolve Holliday structures (2). It has also been shown to recognize cruciforms (2, 3) and loops (6). It may also be involved in very short patch repair (5). Its cleavage characteristics involve it cleaving 3' and within 6 nt from the point of DNA perturbation—causing double-stranded breakage (2, 5). T4 endonuclease VII has been shown to cleave single-base-pair mismatches in model experiments with synthetic oligonucleotides up to ~43 bp (7). This work examines the ability of the enzyme to detect mutations rather than its ability to cleave specific mismatches. Thus, when mutant and wild-type homoduplexes that differ by a single base pair are melted and hybridized, two

heteroduplex species are formed containing two pairs of mismatched bases. The mutation would be detected if any one strand containing one of the four mismatched bases were cleaved. There are four classes of pairs of mismatched bases (type 1, G·A/T·C; type 2, G·T/A·C; type 3, C·C/G·G; type 4, T·T/A·A) and at least two members of each were tested. Only one example showed no cleavage in any of the strands.

MATERIALS AND METHODS

Enzyme and Buffers. T4 endonuclease VII was prepared from an overexpressing *Escherichia coli* K38 transformant containing gene 49 of T4 phage as described by Kosak and Kemper (7). Stock solutions were at 3700 units/μl as determined by Kosak and Kemper (7), where 1 unit is defined as that amount of enzyme that catalyzes degradation of 50% of very fast sedimenting DNA. The reaction buffer used in the assay was prepared as a 10× concentrate (7). The enzyme dilution buffer was prepared as described (7). Assay conditions required between 250 and 3000 units of T4 endonuclease VII depending on the specific DNA being tested. The annealing buffer was prepared as a 2× concentrate (1.2 M NaCl/12 mM Tris·HCl, pH 7.5/14 mM MgCl₂) as described (8). The kinase buffer, 1× TE (pH 8.0), and the formamide/urea loading dye were prepared as described (9).

DNA Preparation. The DNA used in these experiments was amplified by PCR from genomic DNA [β -globin, phenylalanine hydroxylase (PAH), α_1 -antitrypsin, plasmid DNA (21-hydroxylase and the mouse mottled Menkes gene), or cDNA [pyruvate dehydrogenase E1 α subunit (PDH E1 α), dihydropteridine reductase, and the rhodopsin gene]. Each region contained an example of a known mutation except for the mouse mottled Menkes gene and dihydropteridine reductase mutations, which were previously unpublished.

Each DNA sample was prepared by PCR amplification. The β -globin gene (M8, mutation at nt 26 exon 2; M14, mutation at nt -87; M16, sickle mutation; M21, mutation at nt 17 exon 1) was amplified by using primers a and b (10). A larger segment of the β -globin gene (M15), which contains mutations at nt 745, 16, 74, 81, and 666 within IVSII, was amplified by using primers c and d as described (10). The α_1 -antitrypsin gene (M4, mutation at nt 9989) was amplified as described (11). The PDH E1 α gene (M1, F205L; M19, K387fs; M20, S312fs) was amplified by using primers PDH-P and PDH-E as described (12). The PAH gene M5 (homozygous mutation at IVS12 nt 1), M6 (heterozygous mutation at IVS12 nt 1), and M7 (R408W) was PCR amplified by using primers A and B as described (13). The PAH gene (M13, F39L exon 2) was PCR amplified by using the primers 5'-d(GCATCTTATCCTGTA GGA AA)-3' and 5'-d(AGTACTGACCTC AAA TAA GC)-3'. The PCR conditions were 105 s at 95°C, 150 s at 58°C, and 3 min at 72°C for 35 cycles. The 340-bp section of the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PAH, phenylalanine hydroxylase; PDH, pyruvate dehydrogenase; EMC, enzyme mismatch cleavage; CCM, chemical cleavage of mismatch.

Table 1. Summary of mutations tested for by EMC

Type	Sample	Base change	Sequence context	Mismatch set	Mismatch detected	Total fragment length, bp	Nonspecific cleavage†	Fig.
1	M1 (Hom)	C → A	ATT <u>CGAA</u>	A/G/T-C	C-T	797	-	1
1	M2‡ (Hom)	A → C	CCC <u>AATC</u>	A/G/T-C	A-G/T-C	340	-	
2		T → C§	CACITGC	G-T/C-A	ND¶		-	2
2	M3 (Hom)	T → C§	CACITGC	G-T/C-A	G-T	178	-	
2	M4 (Hom)	G → A	GACGAGA	G-T/C-A	C-A	220	+	
2	M5 (Hom)	G → A§	ACAGTAA	G-T/C-A	G-T/C-A	245	+	
2	M6 (Het)	G → A§	ACAGTAA	G-T/C-A	G-T/C-A	245	+	
2	M7 (Het)	C → T	TACCTCG	G-T/C-A	G-T/C-A	245	+	
2	M8 (Hom)	C → T	ACCCAGA	G-T/C-A	G-T	627	-	
2	M9 (Het)	A → G	CCAATGC	G-T/C-A	?	1300	-	
2	M10¶ (Het)	T → C	AGCTCTT	G-T/C-A	G-T	779	-	
2	M11¶ (Hom)	A → G	TGGAGGA	G-T/C-A	?	1502	-	
2	M12¶ (Hom)	C → T	GATCATT	G-T/C-A	?	1502	-	3
3	M13 (Het)	C → G	CTTCTCA	C-C/G-G	C-C**	133	+	
3	M14 (Hom)	C → G	CACCTTA	C-C/G-G	C-C	627	-	
3	M15‡ (Het)	C → G	CAGCTAC	C-C/G-G	C-C/G-G	1377	-	
3		C → G	GACCCTT	C-C/G-G	C-C	1377	-	
1		G → T	GGAGAAG	A-G/T-C	A-G/T-C	1377	-	4
2		C → T	TAACAGG	G-T/C-A	G-T/C-A	1377	-	
2		T → C	TATTTCT	G-T/C-A	G-T/C-A	1377	-	
4	M16 (Het)	A → T	CTGAGGA	A-A/T-T	ND	627	-	
4	M17 (Hom)	T → A	TCAICTG	A-A/T-T	A-A/T-T	204	-	
Del	M18¶ (Hom)	Del 33 bp	TAG-AGG	33-bp loops	?	1502	-	
Del	M19 (Hom)	Del 2 bp	TTT-GTC	AA/TT loops	TT loop**	797	-	
Del	M20 (Hom)	Del 7 bp	GGA-AGT	7-bp loops	?	797	-	
Del	M21 (Het)	Del 1 bp	CTG-GGA	A/T loops	ND	627	-	

Source of mutation was either genomic, plasmid, or cDNA. Samples were either homozygous (Hom) or heterozygous (Het). Details of genes involved are given in the text. Sequence context of the sense strand of normal DNA is shown. Underlined base denotes nucleotide involved in the base change. ND, no cleavage was seen in any one of the four strands containing a mismatched base; ?, cleavage was observed at the mismatches but the strand cleaved could not be determined; Del, deletion (bases given flank the mutation).

†Substantial cleavage was seen in homoduplex control DNA (as well as heteroduplex) but the exact position has not been determined.

‡More than 1 base change was present in the fragment studied.

§Exactly the same mutation has been tested for the respective genes.

¶This mutation at nt 118 was not detected when in the presence of a second mutation at nt 138. However, it was detected in a shorter fragment that did not include the mutation at nt 138 (M3).

‡These mutations were previously unknown.

**Postdigestion end-labeling was performed on these fragments also.

21-hydroxylase B gene (M2, mutations at nt 118 and 138) was amplified by using the primers 5'-d(CTG CTG TGG AAC TGG TGG AA)-3' and 5'-d(ACA GGT AAG TGG CTC AGG TC)-3'. The 178-bp section of the 21-hydroxylase B gene (M3, mutations at nt 118) was amplified by using the primers 5'-d(GCT CTT GAG CTA TAA GTG G)-3' and 5'-d(GGG AGG TCG GGC TGC AGCA)-3'. The 21-hydroxylase A gene (M17, mutation at nt 1004) was amplified by using the primers 5'-d(CTG CAC AGC GGC CTG CTG AA)-3' and 5'-d(CAG TTC AGG ACA AGG AGA GG)-3'. The PCR conditions for the 21-hydroxylase A and B genes were 105 s at 95°C, 150 s at 62°C, and 3 min at 72°C. The dihydropteridine reductase gene (M10, L74P at nt 245) was amplified by using the primers GD and F as described (14). The rhodopsin gene (M9, mutation at codon 15, nt 44) was amplified as described (15). The mouse mottled Menkes gene (M11, mutation at nt 3662; M12, mutation at nt 3367; M18, mutation at nt 4516) was amplified as described (16).

The PCR products were electroeluted onto Whatman I paper in a 1.5% agarose gel and then eluted with 1× TE (pH 8.0). In all cases, the corresponding normal DNA was end-labeled with [γ -³²P]ATP using 5' T4 polynucleotide kinase (Boehringer Mannheim). After the kinase treatment, the DNA was ethanol precipitated and the pellet was washed three times with 70% ethanol to wash away the major portion of unincorporated label. The pellet was resuspended in distilled water to give ~5 ng of end-labeled DNA per μ l.

Heteroduplex formation was performed in 50 μ l (total volume) containing 1× annealing buffer as described (8)

except that the annealing temperature was at 65°C for 1 hr followed by 20 min at room temperature. Calculations of DNA concentration were based on 50–60 ng of unlabeled DNA (10× excess) and 5 ng of end-labeled DNA per single reaction. Heteroduplexes were prepared in bulk in a 50- μ l volume and the pellet was resuspended in the appropriate volume of distilled water. For example, if six reactions were required then 300 ng of mutant DNA and 30 ng of labeled wild-type DNA was used. After the heteroduplex reaction, the pellet was resuspended in 30 μ l of distilled water (i.e., 5 μ l of distilled water is taken per single reaction). An identical procedure was performed in order to prepare labeled homoduplex DNA for the control studies except that an excess amount of unlabeled wild-type DNA was hybridized with the labeled wild-type DNA. This strategy allows two of the four mismatched bases (those present in the labeled strand) to be tested for cleavage by the enzyme.

Enzyme Mismatch Cleavage (EMC). Five microliters of the labeled homoduplex or heteroduplex DNA (50–60 ng) was added to 39 μ l of distilled water and 5 μ l of 10× reaction buffer, all kept on ice. The reaction was initiated by the addition of 1 μ l of the enzyme (100–3000 units/ μ l as specified). The stock solution of enzyme was diluted to the required activity in the enzyme dilution buffer. After addition of the enzyme, the tubes were spun briefly and incubated at 37°C for 1 hr unless otherwise specified. In the case of controls, the enzyme was replaced with 1 μ l of the enzyme dilution buffer and these were incubated. After incubation, the samples were ethanol precipitated, washed in 70% ethanol, dried briefly, and

thoroughly resuspended by vortex mixing in 5 μ l of formamide/urea loading dye. The 5- μ l samples were heated to 100°C and immediately loaded onto an 8% urea/acrylamide sequencing gel. Cleavage products were visualized by autoradiography, and the sizes of the products were compared with radiolabeled ϕ X174 *Hae* III size marker.

Postdigestion End-Labeling. Heteroduplexes were prepared in a 1:1 ratio of unlabeled normal and unlabeled mutant DNA. Heteroduplex formation was identical to that described above except that 25 ng of wild-type DNA and 25 ng of mutant DNA were mixed together and annealed to give 50 ng of duplex DNA per reaction mixture. Fifty nanograms of wild-type DNA was used for the homoduplex control. Enzyme digestion was performed on 50 ng of unlabeled duplex DNA and the products of digestion were end-labeled in a total of 10 μ l of 1 \times kinase buffer, 2 units of 5' polynucleotide kinase, and 1 μ l of a 1:10 dilution of fresh [γ -³²P]ATP. After incubation (45 min at 37°C), the enzyme was denatured at 70°C for 10 min and the reaction mixture was ethanol precipitated. The pellet was washed three times in 70% ethanol, dried briefly, and resuspended in 5 μ l of formamide loading dye.

Chemical Cleavage of Mismatch (CCM). CCM was used as a control method to ensure heteroduplexes had formed and was performed as described (8).

RESULTS

The mutations were chosen to cover at least two examples of each of the four types of single-base-pair mismatch combinations possible when heteroduplexes are formed between mutant and wild-type genes differing by any of the 12 possible single-base changes. Heteroduplex loops resulting from deletions in the mutant DNA were also examined. Results of only one example of each type are given. Results of the other examples are summarized in Table 1. Mutations studied were isolated such that only one mutation occurred in the length studied unless otherwise stated.

Type 1 (Mismatch Set G-A/T-C). M1 contains a homozygous C \rightarrow A mutation 87 bp away from the 5' end of the section of the gene studied (F205L). Using end-labeled wild-type probe, the resulting heteroduplexes contain C*·T and A·G* mismatches. The asterisk designates that the base is present in the labeled strand. After CCM, only the 87-bp fragment was observed on denaturing acrylamide gel since hydroxylamine modifies only the mismatched C. With the EMC, only a single band slightly larger than the 87-bp fragment was observed, suggesting cleavage near the C* in the C*·T mismatch (Fig. 1).

Type 2 (Mismatch Set G-T/A-C). A 245-bp section of the PAH gene (M5) was amplified from genomic DNA of a patient homozygous for a G \rightarrow A mutation (IVS12 nt 1). This mutation occurred 191 bp from the 5' end of the PCR product. End-labeled normal DNA was hybridized with unlabeled mutant DNA. After enzyme cleavage, two bands, one slightly larger than the 191-bp band and another slightly larger than the 54-bp band, were observed resulting from cleavage near the G* in the G*·T mismatch and near the C* in the A·C* mismatch, respectively (Fig. 2).

Type 3 (Mismatch Set C-C/G-G). The PAH mutation (M13) used here involves a heterozygous C \rightarrow G mutation at base 57 in exon 2, 73 bp from the 5' end of the 133-bp section of the gene studied. Heteroduplex formation between labeled normal and unlabeled mutant DNA produced C*·C and G·G* mismatches as well as homoduplexes corresponding to the normal and mutant DNA. CCM using hydroxylamine allowed detection of only the C*·C-containing heteroduplex. Only the 73-bp 5'-end-labeled sense strand derived from the wild-type probe was observed on acrylamide gel electrophoresis. EMC of the same heteroduplex showed a similar pattern except that a band slightly larger than the expected 73-bp fragment was observed (Fig. 3A). This band results from cleavage near the

C*·C heteroduplex. It is clear that the enzyme does not seem to recognize the mismatched G* in the labeled wild-type strand since a 60-bp fragment was not observed. EMC in this section of the PAH gene results in the production of a fragment due to nonspecific cleavage.

Postdigestion end-labeling of this same DNA showed bands slightly larger than the expected 60- and 73-bp bands on the autoradiograph (Fig. 3B), suggesting that there was double-stranded cleavage of the C·C-containing heteroduplex. Cleavage by T4 endonuclease VII results in 3' OH and 5' PO₄ ends. Only the 5' OH ends (primer ends) will label with [γ -³²P]ATP using 5' T4 polynucleotide kinase. Hence, the fragments 3' to the cleavage will NOT be observed on autoradiography. Note that the same nonspecific fragment observed in Fig. 3A is present here also.

Type 4 (Mismatch Set A-A/T-T). A 204-bp section of the 21-hydroxylase A gene (M17) was PCR amplified from plas-

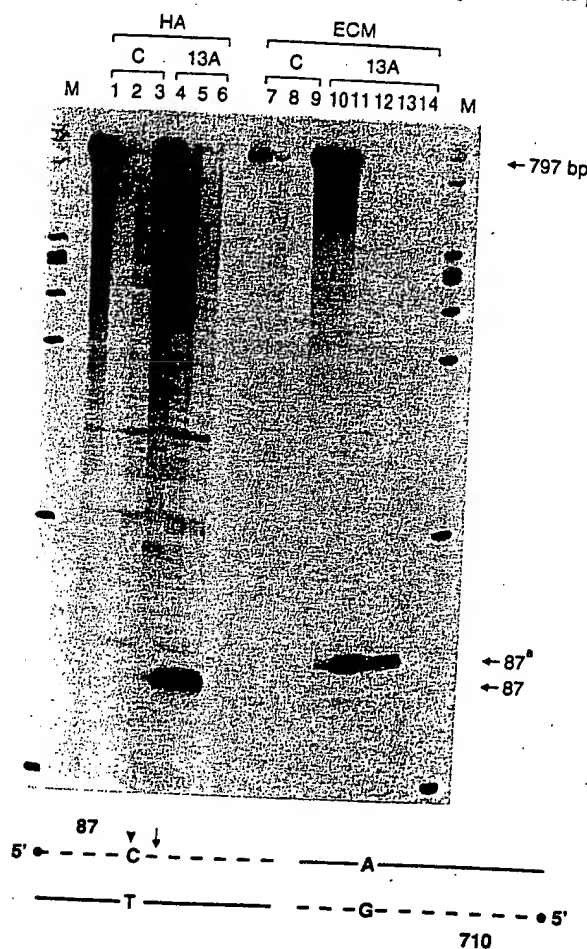


Fig. 1. Autoradiograph of CCM and EMC analysis of the PDH E1 α gene (M1) containing a homozygous C \rightarrow A mutation. Lanes: 1-3, samples of control homoduplex DNA (C) after incubation with hydroxylamine (HA) for 0, 1, and 2 hr; 4-6, samples of test heteroduplex DNA (13A) after incubation with hydroxylamine for 0, 1, and 2 hr; 7-9, samples of C after incubation with 0, 1000, and 3000 units of T4 endonuclease VII; 10-14, samples of 13A after incubation with 0, 250, 1000, 2000, and 3000 units of T4 endonuclease VII. Scheme below represents two types of heteroduplexes that are formed (as described in the text). Broken line, wild-type DNA strands; straight lines, mutant DNA strands; arrows, actual enzyme cleavage sites on the end-labeled strand in each heteroduplex; ●, label; arrowhead, actual cleavage site of the CCM reaction. Superscript a refers to band sizes observed by EMC being slightly larger than the expected band sizes as determined by CCM. Lanes M, end-labeled marker ϕ X174 *Hae* III.

mid DNA. This mutant DNA contained a homozygous T → A mutation at base 1004 of the gene, 110 bp from the 5' end of the section of the gene studied. This results in a heteroduplex species containing both A·A* and T*·T mismatches. EMC on this sample with end-labeled probe showed two bands: one slightly greater than 110 bp and the other slightly greater than 94 bp (Fig. 4). This confirms that the enzyme recognized both strands of the probe. The intensity of the products on autoradiography shows that the enzyme recognizes the A·A* mismatch with greater efficiency than the T*·T mismatch. CCM on this sample showed only the 110-bp fragment obtained after modification and cleavage of the mismatched T base in the sense strand of the probe (Fig. 4).

DISCUSSION

A total of 3 type 1, 13 type 2, 4 type 3, 2 type 4, and 4 deletion mutations have been tested by the enzyme cleavage method in this study. Four of the single-base-pair mutations detected were previously unknown. Of the 18 known single-base mutational changes tested, only 1 (an A → T change) did not show cleavage of any strand of the two heteroduplexes (Table 1). This may be due to a sequence context feature since of the four single loops tested (generated by deletions) only one did not show cleavage of any strand, and that one involved the same base as the single-base mutational change A → T present in M16. We would like to investigate the reasons for nondetection

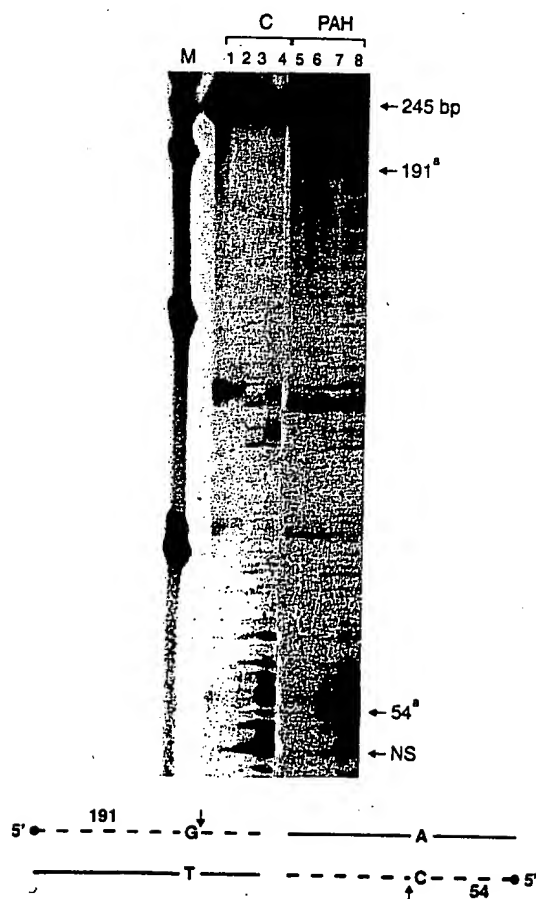


Fig. 2. Autoradiograph of EMC analysis of the PAH gene (M5) containing a homozygous G → A mutation. Lanes: 1–4, samples of control homoduplex DNA (C) after incubation with 0, 250, 500, and 1000 units of T4 endonuclease VII; 5–8, samples of test heteroduplex DNA (PAH) after incubation with 0, 250, 500, and 1000 units of T4 endonuclease VII. Scheme below is as in Fig. 1.

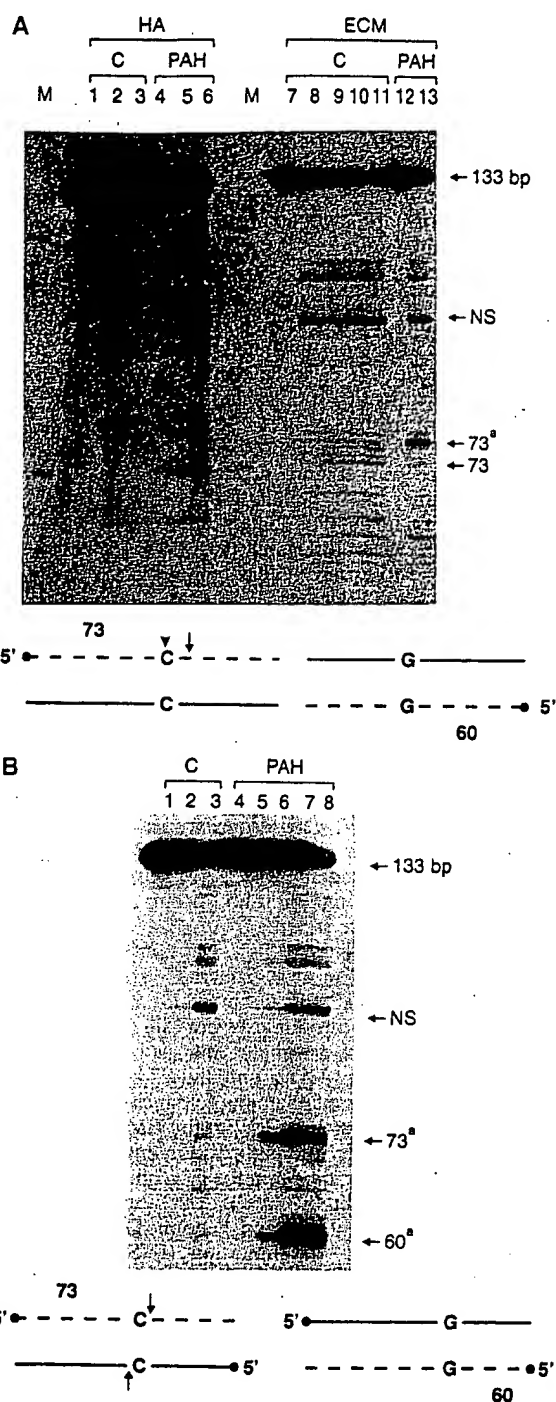


Fig. 3. (A) Autoradiograph of CCM and EMC analysis of the PAH gene (M13) containing a heterozygous C → G mutation in exon 2. Lanes: 1–3, samples of control homoduplex DNA (C) after incubation with hydroxylamine for 0, 1, and 1.5 hr; 4–6, samples of test heteroduplex DNA (PAH) after incubation with hydroxylamine for 0, 1, and 1.5 hr; 7–11, samples of C after incubation with 0, 1000, 2000, 2500, and 3000 units of T4 endonuclease VII; 12 and 13, samples of PAH after incubation with 0 and 1000 units of T4 endonuclease VII. (B) Autoradiograph of postdigestion end-labeling (as described in the text) of the PAH gene (M13) containing a heterozygous C → G mutation in exon 2. Lanes: 1–3, samples of control homoduplex DNA (C) after incubation with 0, 250, and 1000 units of T4 endonuclease VII; 4–8, samples of test heteroduplex DNA (PAH) after incubation with 0, 100, 250, 500, and 1000 units of T4 endonuclease VII. Scheme below is as in Fig. 1.

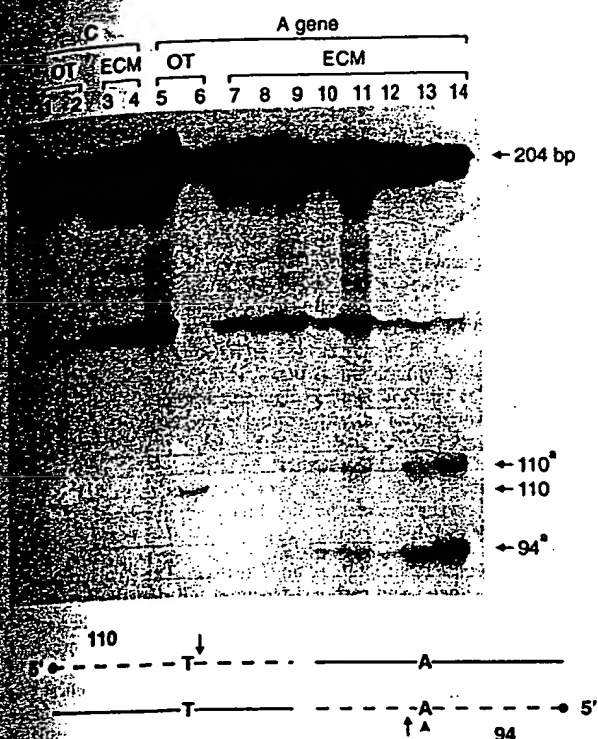


FIG. 4. Autoradiograph of CCM and EMC analysis of the 21-hydroxylase A gene (M17) in a plasmid containing a homozygous T → A mutation. Lanes: 1 and 2, samples of control homoduplex DNA (C) after incubation with osmium tetroxide (OT) for 0 and 5 min; 3 and 4, samples of C after incubation with 0 and 250 units of T4 endonuclease VII; 5 and 6, samples of test heteroduplex DNA (A gene) after incubation with OT for 0 and 5 min; 7 and 8, samples of A gene after incubation with 0 and 250 units of T4 endonuclease VII; 9–11, samples of A gene after incubation with 500 units of T4 endonuclease VII for 1, 3, and 16 hr; 12–14, samples of A gene after incubation with 1000 units of T4 endonuclease VII for 1, 3, and 16 hr at 37°C. Scheme below is as in Fig. 1.

of any mutations and attempt to change conditions appropriately to allow detection. Similarly, we would like to investigate the reasons for nonspecific cleavage present in some cases (see Table 1 and Fig. 3).

Our results show that in about half the cases studied, detection was observed by the cleavage of one of the heteroduplexes in the set. For example, the heteroduplex containing the A-G mismatch was cleaved in M2, but we relied on cleavage of the heteroduplex containing the reciprocal T-C mismatch for detection in M1 (Table 1). This gives further support to the view that T4 endonuclease VII is dependent on sequence context as well as DNA structure (5, 17). It is also clear from Table 1 that the mismatch pair generally considered to be the most thermostable (G-T) is recognized efficiently by T4 endonuclease VII where 8 of 13 type 2 mismatches tested showed cleavage of G-T-containing heteroduplex. In these cases, detection depended either solely (in three cases) or in conjunction with recognition of the complementary C-A mis-

match pair (in five cases). At the other end of the scale, the mismatch pair considered to be one of the least thermostable (C-C) was recognized by T4 endonuclease VII in all four cases tested in this study.

The postdigestion end-labeling method described here was developed to apply these findings to screen lengths of DNA in the most effective manner. Most experiments were performed with excess unlabeled target DNA over labeled probe DNA to form duplexes before cleavage. For simple and practical use, we propose forming duplexes between equimolar mutant and wild-type DNA, cleaving and then kinase labeling all 5' OH ends before electrophoresis. This allows assay of each strand for cleavage without probe production, thus maximizing the chances of detecting mutations. When using this method, two bands were always observed resulting from the labeling of all the free 5' OH ends of the cleavage products.

Drs. Sue Forrest, Henrik Dahl, and Julian Mercer are thanked for useful discussions. Dr. Sue Forrest is thanked for her critical review of this paper. The EMC experiments on the Menkes gene were conducted by Andrew Grimes. The EMC experiment involving the rhodopsin gene was conducted by George Makris. DNA used in the this study was donated by H. Dahl (PDH E1α), S. Forrest (PAH, α₁-antitrypsin), C. Camaschella (β-globin), S. Ramus (PAH), and R. Campbell (21-hydroxylase A gene and B gene). Grant support was from the Australian National Health and Medical Research Council (R.Y. and R.G.H.C.) and the Victorian Health Promotion Foundation (R.Y. and R.G.H.C.).


1. Cotton, R. G. H. (1993) *Mutat. Res.* 285, 125–144.
2. Mizuuchi, K., Kemper, B., Hays, J. & Weisberg, R. A. (1982) *Cell* 29, 357–365.
3. Bhattacharyya, A., Murchie, A. I. H., von Kitzing, E., Diekmann, S., Kemper, B. & Lilley, D. M. J. (1991) *J. Mol. Biol.* 221, 1191–1207.
4. West, S. C. (1992) *Annu. Rev. Biochem.* 61, 603–640.
5. Solaro, P. C., Birkenkamp, K., Pfeiffer, P. & Kemper, B. (1993) *J. Mol. Biol.* 230, 868–877.
6. Kleff, S. E. & Kemper, B. (1988) *EMBO J.* 7, 1527–1535.
7. Kosak, H. G. & Kemper, B. W. (1990) *Eur. J. Biochem.* 194, 779–784.
8. Saleeba, J. A. & Cotton, R. G. H. (1993) *Methods Enzymol.* 217, 288–295.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
10. Dianzani, I., Camaschella, C., Saglio, G., Forrest, S. M., Ramus, S. & Cotton, R. G. H. (1991) *Genomics* 11, 48–53.
11. Forrest, S. M., Dry, P. J. & Cotton, R. G. H. (1992) *Prenatal Diagn.* 12, 133–137.
12. Dahl, H.-H. M., Maragos, C., Brown, R. M., Hansen, L. L. & Brown, G. K. (1990) *Am. J. Hum. Genet.* 47, 286–293.
13. DiLella, A. G., Huang, W. M. & Woo, S. L. C. (1988) *Lancet* 1, 497–499.
14. Smooker, P. M., Howells, D. W. & Cotton, R. G. H. (1990) *Biochemistry* 32, 6443–6449.
15. Sullivan, L. J., Makris, G. S., Dickinson, P., Mulhall, L. E. M., Forrest, S., Cotton, R. G. H. & Loughnan, M. S. (1993) *Arch. Ophthalmol.* 111, 1512–1517.
16. Mercer, J. F. B., Grimes, A., Ambrosini, L., Lockhart, P., Paynter, J. A., Dierick, H. & Glover, T. W. (1994) *Nat. Genet.* 6, 374–378.
17. Pottmeyer, S. & Kemper, B. (1992) *J. Mol. Biol.* 223, 607–615.

GM-13762) and Keiichi Itakura (GM-25658) and funds from NIH biomedical Research Support Grant 5S07 RR05471-16 to the City of Ope National Medical Center. We are grateful to the excellent technical assistance of Merrie Jo Johnson.

REFERENCES

1. Wallace, R.B. (1978) *The Cell Nucleus*; Busch, H., Ed.; Vol. 6, pp. 35-74; Academic Press, New York.
2. Lewin, B. (1975) *Cell* 4, 77-93.
3. Davidson, E.H. (1976) *Gene Activity in Early Development*; Academic Press, New York.
4. Tonegawa, S.; Brack, C.; Hozumi, N. and Schuller, R. (1977) *Proc. Nat. Acad. Sci. USA* 74, 3518-3522.
5. Garapin, A.C.; Lepenne, J.P.; Roskam, W.; Perrin, F.; Cami, B.; Krust, A.; Breathnach, R.; Chambon, P. and Kourilsky, P. (1978) *Nature* 273, 349-354.
6. Maniatis, T.; Haridson, R.C.; Lacy, E.; Lauer, J.; O'Connell, C.; Quon, D.; Sim, G.K. and Efstratiadis, A. (1978) *Cell* 15, 687-701.
7. Montgomery, D.L.; Hall, B.D.; Gillam, S. and Smith, M. (1978) *Cell* 14, 673-680.
8. Sanger, F.; Air, G.M.; Barrel, B.G.; Brown, N.L.; Coulson, A.R.; Fildes, J.C.; Hutchison, C.A.; Slocumbe, P.M. and Smith, M. (1977) *Nature* 264, 687.
9. Hirose, T.; Crea, R. and Itakura, K. (1978) *Tetrahedron Letters*, No. 28, 2449-2452.
0. Razin, A.; Hirose, T.; Itakura, K. and Riggs, A.D. (1978) *Proc. Nat. Acad. Sci. USA* 75, 4268-4271.
1. Razin, A.; Sedat, J.W. and Sinsheimer, R.L. (1970) *J. Mol. Biol.* 53, 251-259.
2. Sgaravelia, V. and Khorana, H.G. (1972) *J. Mol. Biol.* 72, 427-444.
3. Denhardt, D.T. (1966) *Biochem. Biophys. Res. Comm.* 23, 641.
4. Murphy, R.F.; Pearson, W.R. and Bonner, J. (1979) Manuscript submitted to *Nucleic Acids Research*.
5. Pearson, W.R.; Davidson, E.H. and Britten, R.J. (1977) *Nucleic Acid Res.* 4, 1727-1737.
6. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
7. Doel, M.T. and Smith, M. (1973) *FEBS Letters* 34, 99-102.
8. Wu, R. (1972) *Nature New Biol.* 236, 198-200.
9. Padmanabhan, R.; Padmanabhan, R. and Wu, R. (1972) *Biochem. Biophys. Res. Comm.* 48, 1295-1302.
0. Besmer, P.; Miller, R.C.; Caruthers, M.H.; Kumar, A.; Minamoto, K.; van de Sande, J.H.; Sivadara, N. and Khorana, H.G. (1972) *J. Mol. Biol.* 72, 503-522.
1. Britten, R.J.; Graham, D.E. and Neufeld, B.R. (1974) *Methods in Enzymology*; Grossman, L. and Moldave, K., Eds.; Vol. 29, Pt. E, pp. 363; Academic Press.
2. Mettner J.G. and Davidson, N. (1968) *J. Mol. Biol.* 31, 349-370.
3. Uhlenbeck, O.C.; Martin, F.H. and Doty, P. (1971) *J. Mol. Biol.* 57, 217-229.

24. Borer, P.N.; Dengler, B.; Tinoco, I. and Uhlenbeck, O.C. (1974) *J. Mol. Biol.* 86, 843-853.
25. Gilliam, S.; Waterman, K. and Smith, M. (1975) *Nucleic Acid Res.* 2, 625-634.
26. Dodgson, J.B. and Wells, R.D. (1977) *Biochemistry* 16, 2367-2374.

 ExPASy Home page	Site Map	Search ExPASy	Contact us	Swiss-Prot
Hosted by NCSC US Mirror sites: Bolivia China Switzerland Taiwan				

C30

The Canadian and Korean ExPASy sites, ca.expasy.org and kr.expasy.org, are temporarily not available.

Search ☒ for

NiceProt View of Swiss-Prot: P55060

[Printer-friendly view](#)
[Quick BlastP search](#)

[\[General\]](#)
[\[Name and origin\]](#)
[\[References\]](#)
[\[Comments\]](#)
[\[Cross-references\]](#)
[\[Keywords\]](#)
[\[Features\]](#)
[\[Sequence\]](#)
[\[Tools\]](#)

Note: most headings are clickable, even if they don't appear as links. They link to the [user manual](#) or [other documents](#).

General information about the entry

Entry name	CSE1_HUMAN
Primary accession number	P55060
Secondary accession numbers	O75432 Q9H5B7 Q9NTS0 Q9UP98 Q9UP99 Q9UPA0
Entered in Swiss-Prot in	Release 34, October 1996
Sequence was last modified in	Release 41, February 2003
Annotations were last modified in	Release 41, February 2003
Name and origin of the protein	
Protein name	Importin-alpha re-exporter
Synonyms	Chromosome segregation 1-like protein Cellular apoptosis susceptibility protein
Gene name	CSE1L or CAS
From	<i>Homo sapiens</i> (Human) [TaxID: 9606]
Taxonomy	Eukaryota ; Metazoa ; Chordata ; Craniata ; Vertebrata ; Euteleostomi ; Mammalia ; Eutheria ; Primates ; Catarrhini ; Hominidae ; Homo .

References

[1] SEQUENCE FROM NUCLEIC ACID (ISOFORM 1).

TISSUE=[Placenta](#);

MEDLINE=96036098; PubMed=7479798; [[NCBI](#), [ExPASy](#), [EBI](#), [Israel](#), [Japan](#)]

[Brinkmann U.](#), [Brinkmann E.](#), [Gallo M.](#), [Pastan I.](#);

"Cloning and characterization of a cellular apoptosis susceptibility gene, the human homologue to the yeast chromosome segregation gene CSE1.";

[Proc. Natl. Acad. Sci. U.S.A.](#) 92:10427-10431(1995).

[2] SEQUENCE FROM NUCLEIC ACID (ISOFORMS 1; 2 AND 3).

TISSUE=[Brain](#);

MEDLINE=99265971; PubMed=10331944; [[NCBI](#), [ExPASy](#), [EBI](#), [Israel](#), [Japan](#)]

[Brinkmann U.](#), [Brinkmann E.](#), [Bera T.K.](#), [Wellmann A.](#), [Pastan I.](#);

"Tissue-specific alternative splicing of the CSE1L/CAS (cellular apoptosis susceptibility) gene.";

[Genomics](#) 58:41-49(1999).

[3] SEQUENCE FROM NUCLEIC ACID.

MEDLINE=21638749; PubMed=11780052; [[NCBI](#), [ExPASy](#), [EBI](#), [Israel](#), [Japan](#)]

[Deloukas P.](#), [Matthews L.H.](#), [Ashurst J.](#), [Burton J.](#), [Gilbert J.G.R.](#), [Jones M.](#), [Stavrides G.](#), [Almeida J.P.](#), [Babbage A.K.](#), [Bagguley C.L.](#), [Bailey J.](#), [Barlow K.F.](#), [Bates K.N.](#), [Beard L.M.](#), [Beare D.M.](#),

Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dharni P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).

[4] FUNCTION.

MEDLINE=97462907; PubMed=9323134; [NCBI, ExPASy, EBI, Israel, Japan]

Kutay U., Bischoff F.R., Kostka S., Kraft R., Gorlich D.;

"Export of importin-alpha from the nucleus is mediated by a specific nuclear transport factor.";
 Cell 90:1061-1071(1997).

Comments

- **FUNCTION:** Export receptor for importin alpha. Mediates importin-alpha reexport from the nucleus to the cytoplasm after import substrates have been released into the nucleoplasm.
- **SUBUNIT:** Binds with high affinity to importin-alpha only in the presence of RanGTP. The complex is dissociated by the combined action of RanBP1 and RanGAP1.
- **SUBCELLULAR LOCATION:** Nuclear and cytoplasmic.
- **ALTERNATIVE PRODUCTS:**
 - Alternative splicing [3 named forms]

Name 1

Isoform ID P55060-1

This is the isoform sequence displayed in this entry.

Name 2

Isoform ID P55060-2

Features which should be applied to build the isoform sequence: VSP_001222,
VSP_001223.

Name 3

Isoform ID P55060-3

Features which should be applied to build the isoform sequence: VSP_001224,
VSP_001225.

- **TISSUE SPECIFICITY:** HIGHLY EXPRESSED IN PROLIFERATING CELLS.
- **SIMILARITY:** BELONGS TO THE CSE1 FAMILY.
- **SIMILARITY:** Contains 1 importin N-terminal domain.

Copyright

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics

and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

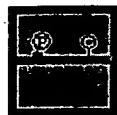
Cross-references

EMBL	U33286; AAC50367.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053640; AAC35007.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053641; AAC35008.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053642; AAC35009.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053651; AAC35297.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053644; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053645; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053646; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053647; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053648; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053649; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AF053650; AAC35297.1; JOINED. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AL121903; CAB86644.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AL121903; CAC33854.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AL133174; CAC14081.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
	AL133174; CAC14082.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
PIR	I39166; I39166.
Genew	HGNC:2431; CSE1L.
CleanEx	HGNC:2431; CSE1L.
MIM	601342 [NCBI / EBI].
GeneCards	CSE1L.
GeneLynx	CSE1L; Homo sapiens.
GO	GO:0005737; Cellular component: cytoplasmic chromosome (<i>traceable author statement</i>).
	GO:0005634; Cellular component: nuclear chromosome (<i>traceable author statement</i>).
	GO:0008262; Molecular function: importin-alpha export receptor activity (<i>traceable author statement</i>).
	GO:0006915; Biological process: apoptosis (<i>traceable author statement</i>).
SOURCE	GO:0008283; Biological process: cell proliferation (<i>traceable author statement</i>).
	CSE1L; Homo sapiens.
Ensembl	P55060; Homo sapiens. [Entry / Contig view]
InterPro	IPR005043; CAS_CSE1.
	IPR001494; Importinb_N.
Pfam	Graphical view of domain structure.
	PF03378; CAS_CSE1; 1.
PROSITE	PF03810; IBN_NT; 1.
	PS50166; IMPORTIN_B_NT; 1.
ProDom	[Domain structure / List of seq. sharing at least 1 domain]
HOVERGEN	[Family / Alignment / Tree]
BLOCKS	P55060.
ProtoNet	P55060.
ProtoMap	P55060.
PRESAGE	P55060.

DIP [P55060](#).
 ModBase [P55060](#).
 SWISS-2DPAGE [Get region on 2D PAGE](#).

Keywords

Transport; Protein transport; Nuclear protein; Alternative splicing.

Features

[Feature table viewer](#)

Key	From	To	Length	Description	FTid
DOMAIN	29	102	74	IMPORTIN N-TERMINAL.	
VARSPLIC	190	195		ATIELC -> VWNASW (in <u>isoform 2</u>).	VSP_001222
VARSPLIC	196	971		Missing (in <u>isoform 2</u>).	VSP_001223
VARSPLIC	943	945		VPS -> TYF (in <u>isoform 3</u>).	VSP_001224
VARSPLIC	946	971		Missing (in <u>isoform 3</u>).	VSP_001225
CONFLICT	231	233		WEG -> FED (IN REF. <u>2</u> ; <u>AAC35297</u> AND <u>3</u>).	
CONFLICT	514	514		G -> E (IN REF. <u>2</u> ; <u>AAC35297</u> AND <u>3</u>).	
CONFLICT	848	848		K -> N (IN REF. <u>1</u>).	
CONFLICT	934	934		K -> M (IN REF. <u>1</u>).	

Sequence information

Length: 971 Molecular weight: 110325 CRC64: 850F2F07B954E316 [This is a checksum on the sequence]
 AA Da

10	20	30	40	50	60
MELSDANLQT	LTEYLKKTLD	PDPAIRRPAE	KFLESVEGNQ	NYPLLLLTLL	EKSQDNVIKV
70	80	90	100	110	120
CASVTFKNYI	KRNWRIVEDE	PNKICEADRV	AIKANIVHLM	LSSPEQIQKQ	LSDAISIIGR
130	140	150	160	170	180
EDFPQKWPD	LTEMVNRFQS	GDFHVGINGVL	RTAHSLEFKRY	RHEFKSNELW	TEIKLVLDLF
190	200	210	220	230	240
ALPLTNLFKA	TIELCSTHAN	DASALRILFS	SLILISKLFY	SLNFQDLPEF	WEGNMETWMN
250	260	270	280	290	300
NFHTLLTLDN	KLLQTDDEEE	AGLLELLKSQ	ICDNAALYAQ	KYDEEFQRYL	PRFVTAIWNL
310	320	330	340	350	360
LVTGQEVKY	DLLVSNAIQF	LASVCERPHY	KNLFEDQNTL	TSICEKVIVP	NMEFRAADEE
370	380	390	400	410	420
AFEDNSEEYI	RRDLEGSDD	TRRRAACDLV	RGLCKFFEGP	VTGIFSGYVN	SMLQEYAKNP
430	440	450	460	470	480

```

SVNWKHKDAA IYLVTSLSK AQTQKHGITQ ANELVNLTEF FVNHILPDLK SANVNEFPVL
    490      500      510      520      530      540
KADGIKYIMI FRNQVPKEHL LVSIPLLINH LQAGSIVVHT YAAHALERLF TMRGPNNATL
    550      560      570      580      590      600
FTAAEIAPFV EILLTNLFKA LTLPGSSENE YIMKAIMRSF SLLQEAIIPY IPTLITQLTQ
    610      620      630      640      650      660
KLLAVSKNPS KPHFNHYMFE AICLSIRITC KANPAAVVNF EEALFLVFTE ILQNDVQEFI
    670      680      690      700      710      720
PYVFQVMSLL LETHKNDIPS SYMALFPHLL QPVLWERTGN IPALVRLQA FLERGSNTIA
    730      740      750      760      770      780
SAAADKIPGL LGVFQKLIAS KANDHQGFYL LNSIIEHMPP ESVDQYRKQI FILLFQRLQN
    790      800      810      820      830      840
SKTTKFIKSF LVFINLYCIK YGALALQEIF DGIQPKMFGM VLEKIIPEI QKVSGNVEKK
    850      860      870      880      890      900
ICAVGITKLL TECPPMDTE YTKLWTPLLQ SLIGLFELPE DDTIPDEEHF IDIEDTPGYQ
    910      920      930      940      950      960
TAFSQLAFAG KKEHDPVGQM VNNPKIHLAQ SLHKLSTACP GRVPSMVSTS LNAEALQYLQ
    970
GYLQAASVTL L

```

P55060 in [FASTA format](#)

[View entry in original Swiss-Prot format](#)

[View entry in raw text format \(no links\)](#)

[Report form for errors/updates in this Swiss-Prot entry](#)

BLAST

[BLAST submission on ExPASy/SIB](#)
or at [NCBI \(USA\)](#)



Sequence analysis tools: [ProtParam](#), [ProtScale](#),
[Compute pI/Mw](#), [PeptideMass](#), [PeptideCutter](#),
[Dotlet](#) (Java)



[ScanProsite](#), [MotifScan](#)



Search the [SWISS-MODEL Repository](#)



[ExPASy Home page](#)

[Site Map](#)

[Search ExPASy](#)

[Contact us](#)

[Swiss-Prot](#)

Hosted by [NCSC US](#) Mirror sites: [Bolivia](#) [China](#) [Switzerland](#) [Taiwan](#)

The Canadian and Korean ExPASy sites, [ca.expasy.org](#) and [kr.expasy.org](#), are temporarily not available.

C30

ID CSE1_HUMAN STANDARD; PRT; 971 AA.
 AC P55060; Q9UP99; O75432; Q9NTS0; Q9H5B7; Q9UP98; Q9UPA0;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE IMPORTIN-ALPHA RE-EXPORTER (CHROMOSOME SEGREGATION 1-LIKE PROTEIN)
 DE (CELLULAR APOPTOSIS SUSCEPTIBILITY PROTEIN).
 GN CSE1L OR CAS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=PLACENTA;
 RX MEDLINE=96036098; PubMed=7479798;
 RA Brinkmann U., Brinkmann E., Gallo M., Pastan I.;
 RT "Cloning and characterization of a cellular apoptosis susceptibility
 RT gene, the human homologue to the yeast chromosome segregation gene
 RT CSE1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:10427-10431(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=BRAIN;
 RX MEDLINE=99265971; PubMed=10331944;
 RA Brinkmann U., Brinkmann E., Bera T.K., Wellmann A., Pastan I.;
 RT "Tissue-specific alternative splicing of the CSE1L/CAS (cellular
 RT apoptosis susceptibility) gene.";
 RL Genomics 58:41-49(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehtvaslaiho M.H., Liversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [4]

RP FUNCTION.

RX MEDLINE=97462907; PubMed=9323134;

RA Kutay U., Bischoff F.R., Kostka S., Kraft R., Gorlich D.;

RT "Export of importin-alpha from the nucleus is mediated by a specific

RT nuclear transport factor.";

RL Cell 90:1061-1071(1997).

CC -!- FUNCTION: EXPORT RECEPTOR FOR IMPORTIN ALPHA. MEDIATES IMPORTIN-

CC ALPHA REEXPORT FROM THE NUCLEUS TO THE CYTOPLASM AFTER IMPORT

CC SUBSTRATES HAVE BEEN RELEASED INTO THE NUCLEOPLASM.

CC -!- SUBUNIT: BINDS WITH HIGH AFFINITY TO IMPORTIN-ALPHA ONLY IN THE

CC PRESENCE OF RANGTP. THE COMPLEX IS DISSOCIATED BY THE COMBINED

CC ACTION OF RANBP1 AND RANGAP1.

CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.

CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.

CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROLIFERATING CELLS.

CC -!- SIMILARITY: BELONGS TO THE CSE1 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U33286; AAC50367.1; -.

DR EMBL; AF053640; AAC35007.1; -.

DR EMBL; AF053641; AAC35008.1; -.

DR EMBL; AF053642; AAC35009.1; -.

DR EMBL; AF053651; AAC35297.1; -.

DR EMBL; AF053644; AAC35297.1; JOINED.

DR EMBL; AF053645; AAC35297.1; JOINED.

DR EMBL; AF053646; AAC35297.1; JOINED.

DR EMBL; AF053647; AAC35297.1; JOINED.

DR EMBL; AF053648; AAC35297.1; JOINED.

DR EMBL; AF053649; AAC35297.1; JOINED.

DR EMBL; AF053650; AAC35297.1; JOINED.

DR EMBL; AL121903; CAB86644.1; -.

DR EMBL; AL121903; CAC33854.1; -.

DR EMBL; AL133174; CAC14081.1; -.

DR EMBL; AL133174; CAC14082.1; -.

DR MIM; 601342; -.

DR InterPro; IPR001494; IBN_NT.

DR Pfam; PF03378; CAS_CSE1; 3.

KW Transport; Protein transport; Nuclear protein; Alternative splicing.

FT	<u>VARSPLIC</u>	190	195	ATIELC -> VWNASW (IN ISOFORM 2).
FT	<u>VARSPLIC</u>	196	971	MISSING (IN ISOFORM 2).
FT	<u>VARSPLIC</u>	943	945	VPS -> TYF (IN ISOFORM 3).
FT	<u>VARSPLIC</u>	946	971	MISSING (IN ISOFORM 3).
FT	<u>CONFLICT</u>	231	233	WEG -> FED (IN REF. 2; AAC35297 AND REF.
FT				3).
FT	<u>CONFLICT</u>	514	514	G -> E (IN REF. 2; AAC35297 AND REF. 3).
FT	<u>CONFLICT</u>	848	848	K -> N (IN REF. 1).
FT	<u>CONFLICT</u>	934	934	K -> M (IN REF. 1).

SQ SEQUENCE 971 AA; 110325 MW; 850F2F07B954E316 CRC64;

MELSDANLQT LTEYLKKTLD PDPAIRPAE KFLESVEGNQ NYPLLLLTLL EKSQDNVIKV

CASVTFKNYI KRNWRIVEDE PNKICEADRV AIKANIVHLM LSSPEQIQKQ LSDAISIIGR

EDFPQKWPDL LTEMVNRFSQ GDFHVGVL RTAHSLEFKRY RHEFKSNELW TEIKLVLDLF

ALPLTNLFKA TIELCSTHAN DASALRILFS SLILISKLFY SLNFQDLPEF WEGNMETWMN

NFHTLLTLDN KLLQTDDEEE AGLLELLKSQ ICDNAALYAQ KYDEEFQRYL PRFVTAIWNL
LVTTGQEVKY DLLVSNAIQF LASVCERPHY KNLFEDQNTL TSICEKVIVP NMEFRAADEE
AFEDNSEEYI RRDLEGS DID TRRRACDLV RGLCKFFEGP VTGIFSGYVN SMLQEYAKNP
SVNWKHKDAA IYLVTS LASK AQTQKHGITQ ANELVNLTEF FVNHILPDLK SANVNEFPVL
KADGIKYIMI FRNQVPKEHL LVSIPLLINH LQAGSIVVHT YAAHALERLF TMRGPNNATL
FTAAEIAPFV EILLTNLFKA LTLPGSSENE YIMKAIMRSF SLLQEAIIPY IPTLITQLTQ
KLLAVSKNPS KPHFNHYMFE AICLSIRITC KANPAAVVNF EEALFLVFTE ILQNDVQEFI
PYVFQVMSLL LETHKNDIPS SYMALFPHLL QPVLWERTGN IPALVRLLQA FLERGSNTIA
SAAADKIPGL LGVFQKLIAS KANDHQGFYL LNSIIHMPP ESVDQYRKQI FILLFQRLQN
SKTTKFIKSF LVFINLYCIK YGALALQEIF DGIQPKMFGM VLEKIIPEI QKVSGNVEKK
ICAVGITKLL TECPPMMDTE YTKLWTPLLQ SLIGLFELPE DDTIPDEEHF IDIEDTPGYQ
TAFSQLAFAG KKEHDPVGQM VNNPKIHLAQ SLHKLSTACP GRVPSMVSTS LNAEALQYLQ
GYLQAASVTL L

//